Second Edition

Principles of Laser Materials Processing

Developments and Applications

Elijah Kannatey-Asibu, Jr.

WILEY

PRINCIPLES OF LASER MATERIALS PROCESSING: DEVELOPMENTS AND APPLICATIONS

Second Edition

Elijah Kannatey-Asibu, Jr.
University of Michigan
Ann Arbor
MI, USA

WILEY

This second edition first published 2023
© 2023 John Wiley & Sons, Inc.

Edition History
John Wiley & Sons, Inc. (1e, 2009)

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by law. Advice on how to obtain permission to reuse material from this title is available at http://www.wiley.com/go/permissions.

The right of Elijah Kannatey-Asibu, Jr., to be identified as the author of this work has been asserted in accordance with law.

Registered Office
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Trademarks: Wiley and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.

Limit of Liability/Disclaimer of Warranty
In view of ongoing research, equipment modifications, changes in governmental regulations, and the constant flow of information relating to the use of experimental reagents, equipment, and devices, the reader is urged to review and evaluate the information provided in the package insert or instructions for each chemical, piece of equipment, reagent, or device for, among other things, any changes in the instructions or indication of usage and for added warnings and precautions. While the publisher and authors have used their best efforts in preparing this work, they make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials or promotional statements for this work. The fact that an organization, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and authors endorse the information or services the organization, website, or product may provide or recommendations it may make. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

Library of Congress Cataloging-in-Publication Data

Names: Kannatey-Asibu, E., author.
Title: Principles of laser materials processing : developments and applications / Elijah Kannatey-Asibu, Jr.
Description: Second edition. | Hoboken, NJ : Wiley, 2023. | Includes bibliographical references and index.
Identifiers: LCCN 2022035658 (print) | LCCN 2022035659 (ebook) | ISBN 9781119881605 (cloth) | ISBN 9781119881612 (adobe pdf) | ISBN 9781119881629 (epub)
Subjects: LCSH: Lasers–Industrial applications. | Materials science.
Classification: LCC TA1675 .K36 2023 (print) | LCC TA1675 (ebook) | DDC 621.36/6–dc23/eng/20221006
LC record available at https://lccn.loc.gov/2022035658
LC ebook record available at https://lccn.loc.gov/2022035659

Cover Design: Wiley
Cover Image: © Aumm graphixphoto/Shutterstock

Set in 10/12pt TimesLTStd by Straive, Chennai, India

To the memory of my parents
Kofi Kannatey and Efuwa Edziiba
And my children
Bianca, Araba, and Kwame

CONTENTS

PREFACE TO THE SECOND EDITION xxi

PREFACE TO THE FIRST EDITION xxiii

ABOUT THE COMPANION WEBSITE xxv

PART I PRINCIPLES OF INDUSTRIAL LASERS 1

1 Laser Background 3

1.1 Laser Generation 3
1.1.1 Atomic Transitions 3
1.1.1.1 Population Distribution 3
1.1.1.2 Absorption 4
1.1.1.3 Spontaneous Emission 5
1.1.1.4 Stimulated Emission 6
1.1.1.5 Einstein Coefficients: A_e, B_{12}, B_{21} 6
1.1.2 Lifetime 8
1.1.3 Optical Absorption 8
1.1.4 Population Inversion 9
1.1.5 Threshold Gain 9
1.1.6 Two-Photon Absorption 10
1.2 Optical Resonators 12
1.2.1 Standing Waves In A Rectangular Cavity 12
1.2.2 Planar Resonators 13
1.2.2.1 Beam Modes 14
1.2.2.2 Line Selection 17
1.2.2.3 Mode Selection 18
1.2.3 Confocal Resonators 18
1.2.4 Concentric Resonators 20
1.3 Laser Pumping 21
1.3.1 Optical Pumping 21
1.3.1.1 Arc or Flashlamp Pumping 21
1.3.1.2 Diode Laser Pumping 22
1.3.2 Electrical Pumping 23

1.4 System Levels 24
1.4.1 Two-Level System 24
1.4.2 Three-Level System 24
1.4.2.1 Four-Level System 25
1.5 Broadening Mechanisms 26
1.5.1 Line Shape Function 26
1.5.2 Line-Broadening Mechanisms 26
1.5.2.1 Homogeneous Broadening 26
1.5.2.2 Inhomogeneous Broadening 28
1.5.3 Comparison of Individual Mechanisms 29
1.6 Beam Modification 29
1.6.1 Quality Factor 30
1.6.2 Q-Switching 31
1.6.3 Mode Locking 33
1.7 Beam Characteristics 35
1.7.1 Beam Divergence 36
1.7.2 Monochromaticity 37
1.7.3 Beam Coherence 38
1.7.3.1 Spatial Coherence 38
1.7.3.2 Temporal Coherence 40
1.7.4 Intensity and Brightness 42
1.7.5 Focusing 42
1.8 Summary 43
Appendix 1.A 44
Problems 46
Bibliography 53

2 Types of Lasers 55

2.1 Solid-State Lasers 55
2.1.1 The Nd:YAG Laser 57
2.1.2 The Nd:Glass Laser 57
2.2 Gas Lasers 57
2.2.1 Neutral Atom Lasers 58
2.2.2 Ion Lasers 60
2.2.3 Molecular Gas Lasers 61
2.2.3.1 Vibrational-Rotational Lasers 62
2.2.3.2 Excimer Lasers 67
2.3 Semiconductor (Diode) Lasers 69
2.3.1 Semiconductor Background 69
2.3.2 Semiconductor Lasers 72
2.3.3 Semiconductor Laser Types 76
2.3.3.1 Homojunction Lasers 76

2.3.3.2 Heterojunction Lasers 76
2.3.3.3 Quantum Well Lasers 77
2.3.4 Low-Power Diode Lasers 78
2.3.5 High-Power Diode Lasers 78
2.3.6 Applications of High-Power Diode Lasers 79
2.4 New Developments in Industrial Laser Technology 80
2.4.1 Slab Lasers 80
2.4.2 Disk Lasers 81
2.4.3 Ultrafast (Femtosecond) Lasers 82
2.4.4 Fiber Lasers 84
2.4.4.1 Pumping 85
2.4.4.2 Pulsed Output 86
2.4.4.3 CW Output 87
2.4.4.4 Power Scaling 87
2.4.4.5 Applications of Fiber Lasers 88
2.4.4.6 Advantages of Fiber Lasers 88
2.4.4.7 Disadvantages of Fiber Lasers 89
2.5 Summary 89
Appendix 2.A 90
Appendix 2.B 91
Appendix 2.C 92
Problems 92
Bibliography 93

3 Beam Delivery 95

3.1 The Electromagnetic Spectrum 95
3.2 Birefringence 96
3.3 Brewster Angle 96
3.4 Polarization 98
3.5 Beam Expanders 101
3.6 Beam Splitters 102
3.7 Beam Delivery Systems 103
3.7.1 Conventional Beam Delivery 103
3.7.2 Fiber Optic Systems 105
3.7.2.1 Optical Fiber Characteristics 105
3.7.2.2 Waveguide Structure 106
3.7.2.3 Background 106
3.7.2.4 Fiber Types 107
3.7.2.5 Beam Degradation 109
3.7.2.6 Application of Optical Fibers in High-Power Laser Systems 116

3.8 Beam Shaping 116
3.8.1 Beam Shaping Using Diffractive Optics 117
3.8.1.1 Fresnel Phase Plate 118
3.8.1.2 Diffractive Optics Design 119
3.8.1.3 Beam Propagation After Diffraction 122
3.8.1.4 Diffractive Optical Element Construction 123
3.8.2 Beam Shaping Using Coherent Beam Combining and Optical Phase Array 124
3.9 Summary 125
Appendix 3.A 126
Problems 128
Bibliography 130

PART II ENGINEERING BACKGROUND 133

4 Heat and Fluid Flow 135

4.1 Energy Balance During Processing 135
4.2 Heat Flow in the Workpiece 136
4.2.1 Temperature Distribution 136
4.2.1.1 Thick Plate with Point Heat Source (Three-dimensional) 139
4.2.1.2 Thin Plate with Line Heat Source (Two-dimensional) 140
4.2.2 Peak Temperatures 146
4.2.3 Cooling Rates 148
4.2.4 Gaussian Heat Source 151
4.2.5 The Two-Temperature Model 152
4.3 Fluid Flow in Molten Pool 156
4.3.1 Continuity Equation 156
4.3.2 Navier–Stokes Equations 157
4.3.3 Surface Tension Effect 158
4.3.4 Free Surface Modeling 160
4.4 Summary 161
Appendix 4.A 162
Appendix 4.B Derivation of Equation (4.2a) 162
Appendix 4.C Moving Heat Source 163
Appendix 4.D 164
Appendix 4.E 165
Appendix 4.F 166
Appendix 4.G 167
Problems 169
Bibliography 172

5 The Microstructure 175

5.1 Process Microstructure 175
5.1.1 Fusion Zone 176
5.1.1.1 Initial Solidification 176
5.1.1.2 Microstructure 182
5.1.1.3 Nucleation and Grain Refinement in Molten Pool 186
5.1.2 Zone of Partial Melting 188
5.1.3 Heat-Affected Zone 188
5.1.3.1 Pure Metals 188
5.1.3.2 Precipitation-Hardening and Nonferrous Alloys 191
5.1.3.3 Steels 191
5.2 Discontinuities 195
5.2.1 Porosity 195
5.2.2 Cracking 197
5.2.2.1 Hot Cracking 198
5.2.2.2 Liquation Cracking 199
5.2.2.3 Cold Cracking 200
5.2.3 Lack of Fusion 201
5.2.4 Incomplete Penetration 201
5.2.5 Undercut 202
5.3 Summary 202
Appendix 5.A 203
Problems 204
Bibliography 207

6 Solidification 209

6.1 Solidification Without Flow 209
6.1.1 Solidification of a Pure Metal 209
6.1.2 Solidification of a Binary Alloy 211
6.1.2.1 Temperature and Concentration Variation in a Solidifying Alloy 211
6.1.2.2 Interface Stability Theories 212
6.1.2.3 Mushy Zone 215
6.2 Solidification with Flow 216
6.2.1 Mushy Fluid 218
6.2.2 Columnar Dendritic Structure 219
6.3 Rapid Solidification 221
6.4 Summary 222
Appendix 6.A 223
Problems 223
Bibliography 224

7 Residual Stresses and Distortion 227

7.1 Causes of Residual Stresses 227
7.1.1 Thermal Stresses 227
7.1.2 Nonuniform Plastic Deformation 230
7.2 Basic Stress Analysis 232
7.2.1 Stress–Strain Relations 233
7.2.1.1 Linear Elastic Behavior 233
7.2.1.2 Plastic Flow of Metals 234
7.2.2 Plane Stress and Plane Strain 235
7.2.2.1 Plane Stress 235
7.2.2.2 Plane Strain 235
7.2.2.3 Plane Stress/Plane Strain Equations 236
7.2.2.4 Compatibility Equation 236
7.2.2.5 Stress–Strain Relations for Plane Stress/Plane Strain 237
7.3 Effects of Residual Stresses 237
7.3.1 Apparent Change in Strength 237
7.3.2 Distortion 238
7.4 Measurement of Residual Stresses 240
7.4.1 Stress Relaxation Techniques 240
7.4.1.1 Sectioning Technique 241
7.4.1.2 Drilling Technique 241
7.4.1.3 Strain Analysis 243
7.4.2 X-ray Diffraction Technique 245
7.4.2.1 Principle of the X-ray Diffraction Technique 245
7.4.2.2 The Film Technique 246
7.4.2.3 The Diffractometer Technique 247
7.4.3 Neutron Diffraction Technique 247
7.4.4 Residual Stress Equilibrium 249
7.5 Relief of Residual Stresses and Distortion 250
7.5.1 Thermal Treatments 250
7.5.1.1 Preheating 250
7.5.1.2 Postheating 251
7.5.1.3 Limitations of Thermal Stress Relief 251
7.5.2 Mechanical Treatments 251
7.5.2.1 Peening 251
7.5.2.2 Proof Stressing 252
7.5.2.3 Vibratory Stress Relief 252
7.6 Summary 252
Appendix 7.A 253
Appendix 7.B 253
Problems 254
Bibliography 259

PART III LASER MATERIALS PROCESSING 261

8 Background on Laser Processing 263

8.1 System-Related Parameters 263
8.1.1 Power and Power Density 263
8.1.2 Wavelength and Focusing 264
8.1.3 Beam Mode 266
8.1.4 Beam Form 267
8.1.5 Beam Quality 267
8.1.6 Beam Absorption 269
8.1.6.1 Measurement of Absorptivity 270
8.1.7 Beam Alignment 271
8.1.8 Motion Unit 272
8.2 Process Efficiency 272
8.3 Disturbances That Affect Process Quality 274
8.4 General Advantages and Disadvantages of Laser Processing 275
8.4.1 Advantages 275
8.4.2 Disadvantages 275
8.5 Summary 275
Appendix 8.A 276
Problems 276
Bibliography 278

9 Laser Cutting and Drilling 279

9.1 Laser Cutting 279
9.1.1 Forms of Laser Cutting 279
9.1.1.1 Fusion Cutting 280
9.1.1.2 Sublimation Cutting 280
9.1.1.3 Photochemical Ablation 280
9.1.2 Components of a Laser Cutting System 281
9.1.3 Processing Conditions 282
9.1.3.1 Beam Power 282
9.1.3.2 Beam Characteristics 282
9.1.3.3 Traverse Speed 284
9.1.3.4 Assist Gas Functions 284
9.1.3.5 Effect of Focal Position 288
9.1.4 Laser Cutting Principles 288
9.1.4.1 Beam Absorption During Laser Cutting 289
9.1.4.2 Process Modeling 291
9.1.5 Quality of Cut Part 296
9.1.5.1 Striations of the Cut Surface 296
9.1.5.2 Dross Formation 297

9.1.6 Material Considerations 299
9.1.6.1 Metals 299
9.1.6.2 Nonmetals 302
9.1.7 Advantages and Disadvantages of Laser Cutting 305
9.1.7.1 Advantages 305
9.1.7.2 Disadvantages 305
9.1.8 Specific Comparison with Conventional Processes 306
9.1.8.1 Laser, Plasma Arc, and Oxy-Acetylene (Oxy-Fuel) Cutting 306
9.1.8.2 Laser Cutting and Electrical Discharge Machining (EDM) 306
9.1.8.3 Laser Cutting and Abrasive Waterjet Machining 306
9.1.8.4 Laser Cutting and Punching/Nibbling 307
9.1.9 Special Techniques 307
9.2 Laser Drilling 308
9.2.1 Forms of Laser Drilling 309
9.2.1.1 Single-Pulse Drilling 309
9.2.1.2 Multi-Pulse Percussion Drilling 310
9.2.1.3 Trepanning 310
9.2.2 Process Parameters 310
9.2.2.1 Beam Characteristics 310
9.2.2.2 Drilling Characteristics 311
9.2.2.3 Process Defects 311
9.2.3 Analysis of Material Removal During Drilling 311
9.2.3.1 Basic Analysis 313
9.2.3.2 Approximate Analysis 314
9.2.4 Advantages and Disadvantages of Laser Drilling 317
9.2.4.1 Advantages 317
9.2.4.2 Disadvantages 318
9.2.5 Applications 318
9.3 New Developments 318
9.3.1 Micromachining 318
9.3.1.1 Transparent Dielectric Materials 319
9.3.1.2 Metals and Semiconductors 320
9.3.1.3 Micro-Explosions 322
9.3.1.4 Micromachining Applications 323
9.3.2 Laser-Assisted Machining 323
9.4 Summary 326
Appendix 9.A 327
Problems 328
Bibliography 330

10 Laser Welding 335

10.1 Laser Welding Parameters 335
10.1.1 Beam Power and Traverse Speed 336

10.1.2 Effect of Beam Characteristics 336
10.1.2.1 Beam Mode 336
10.1.2.2 Beam Stability 337
10.1.2.3 Beam Polarization 337
10.1.2.4 Pulsed Beams 338
10.1.3 Plasma Formation, Gas Shielding, and Effect of Ambient Pressure 339
10.1.3.1 Plasma Formation 339
10.1.3.2 Gas Shielding 340
10.1.3.3 Effect of Ambient Pressure 342
10.1.4 Beam Size and Focal Point Location 342
10.1.5 Joint Configuration 342
10.2 Welding Efficiency 344
10.3 Mechanism of Laser Welding 344
10.3.1 Conduction Mode Welding 345
10.3.2 Keyhole Welding 345
10.3.2.1 Power Absorption in the Keyhole 348
10.3.2.2 Keyhole Characteristics 349
10.4 Material Considerations 355
10.4.1 Steels 355
10.4.2 Nonferrous Alloys 356
10.4.3 Ceramic Materials 357
10.4.4 Dissimilar Metals 358
10.5 Weldment Discontinuities 359
10.5.1 Porosity 359
10.5.2 Humping 359
10.6 Advantages and Disadvantages of Laser Welding 360
10.6.1 Advantages 360
10.6.2 Disadvantages 360
10.7 Special Techniques 360
10.7.1 Multiple-Beam Welding 360
10.7.1.1 Multiple-Beam Preheating and Postheating 361
10.7.1.2 Multiple-Beam Flow Control 365
10.7.2 Arc-Augmented Laser Welding 367
10.7.3 Wobble Welding 367
10.7.4 Remote Laser Welding 369
10.8 Specific Applications 371
10.8.1 Microwelding 371
10.8.2 Laser-Welded Tailored Blanks 371
10.8.2.1 Advantages of Tailored Blank Welding 372
10.8.2.2 Disadvantages of Tailored Blank Welding 372
10.8.2.3 Applications of Laser-Welded Tailored Blanks 373
10.8.2.4 Formability of Tailor-Welded Blanks 375
10.8.2.5 Limiting Thickness or Strength Ratio 376

10.8.3 Laser Transmission Welding of Plastics 379
10.8.4 Laser Brazing 381
10.8.4.1 Non-Autogenous Laser Brazing 381
10.8.4.2 Autogenous Laser Brazing 381
10.9 Summary 382
Appendix 10.A 383
Problems 384
Bibliography 387

11 Laser Surface Modification 391

11.1 Laser Surface Heat Treatment 391
11.1.1 Important Criteria 392
11.1.2 Key Process Parameters 393
11.1.2.1 Beam Power, Size, Speed, and Shielding Gas 393
11.1.2.2 Beam Mode 393
11.1.2.3 Beam Absorption 397
11.1.2.4 Initial Workpiece Microstructure 400
11.1.3 Temperature Field 401
11.1.4 Microstructural Changes in Steels 401
11.1.4.1 Pearlite Dissolution 402
11.1.4.2 Austenite Homogenization 404
11.1.4.3 Transformation to Martensite 406
11.1.5 Nonferrous Alloys 407
11.1.5.1 Solution Treatment 408
11.1.5.2 Aging 408
11.1.6 Hardness Variation 408
11.1.7 Residual Stresses 411
11.1.8 Semiconductors 411
11.1.9 Polymers 412
11.1.10 Advantages and Disadvantages of Laser Surface Treatment 412
11.1.10.1 Advantages 412
11.1.10.2 Disadvantages 413
11.2 Laser Surface Melting 413
11.3 Laser Direct Metal Deposition 414
11.3.1 Processing Parameters 414
11.3.2 Methods for Depositing the Material 414
11.3.2.1 Preplaced Powder 415
11.3.2.2 Presprayed Surfacing Powder 417
11.3.2.3 Prepositioned Chip 417
11.3.2.4 Wire Feed 417
11.3.2.5 Powder Feeding 417
11.3.3 Dilution 418

11.3.4 Advantages and Disadvantages of Laser Deposition 419
11.3.4.1 Advantages 419
11.3.4.2 Disadvantages 419
11.4 Laser Physical Vapor Deposition (LPVD) 419
11.5 Laser Shock Peening 420
11.5.1 Background Analysis 423
11.5.2 Thermal Relaxation at High Temperatures 426
11.5.3 Advantages and Disadvantages of Laser Shock Peening 426
11.5.3.1 Advantages 426
11.5.3.2 Disadvantages 427
11.5.4 Applications 427
11.6 Laser Texturing 427
11.7 Summary 429
Appendix 11.A 430
Appendix 11.B 431
Problems 432
Bibliography 433

12 Laser Forming 437

12.1 Principle of Laser Forming 437
12.2 Process Parameters 439
12.3 Laser-Forming Mechanisms 439
12.3.1 Temperature Gradient Mechanism 440
12.3.2 Buckling Mechanism 441
12.3.3 Upsetting Mechanism 441
12.3.4 Summary of the Forming Mechanisms 442
12.4 Process Analysis 443
12.5 Advantages and Disadvantages 447
12.5.1 Advantages 447
12.5.2 Disadvantages 448
12.6 Applications 448
12.7 Summary 448
Appendix 12.A 449
Problems 449
Bibliography 450

13 Additive Manufacturing 453

13.1 Computer-Aided Design 453
13.1.1 Curve and Surface Design 454
13.1.1.1 Splines 454
13.1.1.2 Bezier Curves 456
13.1.1.3 Surface Representation 457

13.1.2 Solid Modeling 458
13.1.2.1 Constructive Solid Geometry (CSG) 459
13.1.2.2 Boundary Representation (BREP) 459
13.1.3 Software Formats 460
13.1.3.1 The STL Format 460
13.1.3.2 The IGES Format 461
13.1.4 Supports for Part Building 461
13.1.5 Slicing 461
13.2 Part Building 462
13.2.1 Liquid-Based Systems 463
13.2.1.1 Beam Scanning 464
13.2.1.2 Parallel Processing 468
13.2.1.3 Two-Photon Polymerization 468
13.2.2 Powder-Based Systems 471
13.2.2.1 Selective Laser Sintering (SLS) 472
13.2.2.2 Direct Metal Deposition (DMD) 473
13.2.2.3 Binder Jetting 473
13.2.3 Solid-Based Systems 474
13.2.3.1 Fused Deposition Modeling 474
13.2.3.2 Laminated Object Manufacturing 475
13.2.3.3 Wire Deposition 477
13.2.4 Qualitative Comparison of Some Major Systems 477
13.3 Post-Processing 477
13.4 Applications 478
13.4.1 Design 478
13.4.2 Engineering, Analysis, and Planning 479
13.4.3 Manufacturing and Tooling 479
13.4.4 Personalized Production 479
13.5 Advantages and Disadvantages 480
13.5.1 Advantages 480
13.5.2 Disadvantages 480
13.6 Summary 480
Appendix 13.A 481
Problems 481
Bibliography 483

14 Medical and Nanotechnology Applications of Lasers 485

14.1 Medical Applications 485
14.1.1 Medical Devices 485
14.1.2 Therapeutic Applications 488
14.1.2.1 Surgical Procedures 488
14.1.2.2 Opthalmology 488
14.1.2.3 Dermatology 489
14.1.2.4 Dentistry 489

14.2 Nanotechnology Applications 490
14.2.1 Nanoholes and Grating 491
14.2.2 Nanobumps 491
14.2.3 Laser-Assisted Nanoimprint Lithography 493
14.3 Summary 494
Bibliography 494

15 Sensors for Process Monitoring 497

15.1 Laser Beam Monitoring 497
15.1.1 Beam Power 498
15.1.1.1 Pyroelectric or Thermopile Detector 498
15.1.1.2 Beam Dump 498
15.1.2 Beam Mode 499
15.1.2.1 Mechanical Methods 499
15.1.2.2 Camera-Based Methods 501
15.1.2.3 Approximate Methods 503
15.1.3 Beam Size 503
15.1.3.1 Kapton Film 503
15.1.3.2 Other Methods 504
15.2 Process Monitoring 504
15.2.1 Acoustic Emission (AE) 504
15.2.1.1 AE Detection 504
15.2.1.2 Background 506
15.2.1.3 AE Transmission 507
15.2.1.4 Traditional AE Signal Analysis 507
15.2.2 Acoustic Mirror 507
15.2.3 Audible Sound (AS) Emission 509
15.2.4 Infrared/Ultraviolet (IR/UV) Detection Techniques 511
15.2.4.1 Infrared Detection 511
15.2.4.2 Ultraviolet Detection 516
15.2.5 Optical (Vision) Sensing 517
15.2.5.1 Optical Detectors 517
15.2.5.2 Detector Setup 518
15.2.5.3 Edge Detection Methodology 518
15.3 Summary 522
Appendix 15.A 522
Problems 523
Bibliography 524

16 Processing of Sensor Outputs 527

16.1 Signal Transformation 527
16.1.1 The Fourier Transform 527
16.1.2 The Discrete Fourier Transform (DFT) 528

16.1.3 Pitfalls of Digital Analysis 529
16.1.4 The Sampling Theorem 531
16.1.5 Aliasing 531
16.1.6 Leakage 532
16.2 Data Reduction 532
16.2.1 Variance Criterion 532
16.2.2 Fisher Criterion 532
16.3 Pattern Classification 534
16.3.1 Pattern Recognition 534
16.3.1.1 Bayes Decision Theory 534
16.3.1.2 Bayes Decision Rule for Minimum Error 535
16.3.1.3 Discriminant Function Analysis 536
16.3.1.4 Least-Squares Minimum Distance Classification 537
16.3.1.5 System Training 539
16.3.2 Neural Network Analysis 540
16.3.2.1 Standard Neural Networks 540
16.3.2.2 Deep Learning Networks 544
16.3.3 Sensor Fusion 546
16.3.4 Time–Frequency Analysis 546
16.3.4.1 Short-Time Fourier Transform 547
16.3.4.2 Wavelet Transforms 548
16.3.5 Applications in Manufacturing 549
16.4 Summary 550
Appendix 16.A 550
Problems 552
Bibliography 554

17 Laser Safety 557

17.1 Laser Hazards 557
17.1.1 Radiation-Related Hazards 557
17.1.1.1 Mechanisms of Laser Damage 558
17.1.1.2 Major Hazards 558
17.1.2 Nonbeam Hazards 560
17.1.2.1 Electrical Hazards 561
17.1.2.2 Chemical Hazards 561
17.1.2.3 Environmental Hazards 562
17.1.2.4 Fire Hazards 562
17.1.2.5 Explosion Hazards and Compressed Gases 562
17.1.2.6 Other Hazards 562
17.2 Laser Classification 562
17.3 Preventing Laser Accidents 563
17.3.1 Laser Safety Officer (LSO) 563

17.3.2 Engineering Controls 563
17.3.3 Administrative and Procedural Controls 564
17.3.4 Protective Equipment 564
17.3.4.1 Protective Eyewear 564
17.3.4.2 Other Protective Equipment 567
17.3.5 Warning Signs and Labels 569
17.4 Summary 569
Appendix 17.A 571
Problem 572
Bibliography 572

Index **573**

PREFACE TO THE SECOND EDITION

The second edition emphasizes new developments since the first edition was published. Major additions include a more detailed treatment of fiber lasers, diffractive optics, beam shaping, additive manufacturing, and deep learning networks. More specifically, the following additions or changes have been made:

1. Chapters 1 to 7 have been condensed into Chapter 1 since the material is extensively covered in current books on laser physics.
2. Discussion on Applications of High-Power Diode Lasers in Section 8.4.5 (now 2.4.6) is expanded to include blue lasers.
3. Fiber Lasers are given a more detailed treatment in Section 2.5.4.
4. New material on Beam Shaping Using Diffractive Optics as well as the Coherent Beam Combining with Optical Phase Array technique is introduced in Section 3.8.
5. Subsection 7.5.1.3 has been added on Limitations of Thermal Stress Relief.
6. A paragraph on carbon fiber-reinforced plastic (CFRP) composites has been added to Section 9.1.6.2 on Nonmetals.
7. Micro-Explosions are extensively discussed in Section 9.3.1.3.
8. A brief paragraph on Micromachining Applications is added to Section 9.3.1.4.
9. Additional material on laser welding of galvanized steel is added to Section 10.4.1 on Steels, where a new method for eliminating porosity in lap welding of galvanized steel is discussed.
10. In Section 10.4.2 on Non-Ferrous Alloys, a paragraph is added on laser welding of aluminum for the casing of lithium ion cells used in electric vehicles.
11. Section 10.4.4 on Dissimilar Metals is expanded to include a more detailed discussion on the effect of composition differences between the welded materials, using copper and aluminum as an example.
12. In Section 10.7 on Special Techniques, two major subsections are introduced: 10.7.3 on Wobble Welding and 10.7.4 on Remote Laser Welding.
13. Two new sections, 10.8.3 on Laser Transmission Welding of Plastics and 10.8.4 on Laser Brazing, are added, with laser brazing covering both autogenous and non-autogenous processes.
14. In Section 11.1 on Laser Surface Heat Treatment, two major subsections, 11.1.8 on Semiconductors and 11.1.9 on Polymers, are added.
15. Additional material on the use of femtosecond pulse lasers for laser shock peening is added to Section 11.5.
16. Section 11.5.1 on Background Analysis of laser shock peening is expanded.
17. A new section (11.5.2) on Thermal Relaxation at High Temperatures is added.
18. A brief section (11.5.4) on Applications of Laser Shock Peening is added.
19. A major section (11.6) on Laser Surface Texturing is added.

20. Chapter 13, originally titled "Rapid Prototyping," has been retitled "Additive Manufacturing" with new material. These include Section 13.2.2.2 on Direct Metal Deposition (DMD), Section 13.2.3.3 on Wire Deposition, Section 13.4.4 on Personalized Production, and Section 13.5 on Advantages and Disadvantages. Manufacturers of equipment for the different processes have also been updated. Two-photon polymerization, which was originally in Chapter 20 (now 14), has been moved to this chapter, since it involves building components on a micro-scale.
21. Measurement of beam mode, Section 15.1.2, has been organized into three categories that include new subsections on camera-based and knife-edge methods.
22. Section 16.3.2 on Neural Network Analysis has been expanded to include a new subsection (16.3.2.2) on Deep Learning Networks. Section 16.3.2.1 on Standard Neural Networks has also been expanded.

Several new references and problems have been added to a number of chapters.

Professor Emeritus of Mechanical Engineering
University of Michigan
Ann Arbor

Elijah Kannatey-Asibu, Jr.

PREFACE TO THE FIRST EDITION

Application of lasers in materials processing has been evolving since the development (first demonstration) of the laser in 1960. The early applications focused on processes such as welding, machining, and heat treatment. Newer processes that have evolved over the years include laser forming, shock peening, micromachining, rapid prototyping, and nano processing. The book provides a state-of-the-art compilation of material in the major application areas and is designed to provide the background needed by graduate students to prepare them for industry; researchers to initiate a research program in any of these areas; and practicing engineers to update themselves and gain additional insight into the latest developments in this rapidly evolving field.

The book is partitioned into three parts. The first part, **Principles of Industrial Lasers** (Chapters 1 to 9), introduces the reader to basic concepts on the characteristics of lasers, design of their components, and beam delivery. It is presented in a simple enough format that an engineering student without any prior knowledge of lasers can fully comprehend. It helps the reader acquire a basic understanding of how a laser beam is generated, its basic properties, propagation of the beam, and the various types of lasers available and their specific characteristics. Such knowledge is useful to all engineering students, irrespective of their specific interests or area of application. It will enable them to select an appropriate laser for a given application and help them to determine how best to utilize the laser. The coverage starts with a discussion on laser generation – basic atomic structure and how it leads to atomic transitions (absorption, spontaneous emission, and stimulated emission). The concepts of population inversion and gain criterion for laser action are introduced. Optical resonators (planar and spherical) are discussed in relation to beam modes (longitudinal and transverse) and stability of optical resonators. Techniques for line and mode selection are outlined. Various pumping techniques that can be used to achieve inversion are then presented, including more recent developments such as diode pumping. The rate equations are then introduced to provide some insight into the conditions necessary for achieving population inversion for both 3- and 4-level systems. This is followed by a discussion on broadening mechanisms that are responsible for the spread of a laser frequency over a finite range. These include both homogeneous (natural and collision broadening) and inhomogeneous (Doppler) broadening. Beam modification mechanisms such as Q-switching and mode-locking are presented. Having obtained a fundamental background on laser generation, the characteristics of beams that have a more direct impact on their application are then discussed. These include beam characteristics such as divergence, monochromaticity (with reference to broadening), coherence, polarization, intensity and brightness, frequency stabilization, and focusing. Different types of lasers are then discussed with specific emphasis on high-power lasers used in industrial manufacturing. These include solid-state lasers (Nd:YAG and Nd:Glass); gas lasers (neutral atom, ion, metal vapor (copper-vapor) and molecular lasers (CO_2 and excimer); dye lasers; and semiconductor (diode) lasers. Finally, beam delivery systems are introduced, discussing concepts such as the Brewster angle, polarization, beam expanders, beam splitters, and transmissive, reflective, and fiber optics.

The second part, **Engineering Background** (Chapters 10 to 13), reviews the engineering concepts that are needed to analyze the different processes. Topics that are discussed include thermal analysis

and fluid flow, the microstructure that results from the heat effect, solidification of the molten metal for processes that involve melting, and residual stresses that evolve during the process.

The third part, **Laser Materials Processing** (Chapters 14 to 23), provides a more rigorous and detailed coverage on the subject of laser materials processing and discusses the principal application areas such as laser cutting and drilling, welding, surface modification, laser forming, rapid prototyping, sensor systems, process monitoring, and medical and nanoapplications. Sensors that are normally used for monitoring process quality are also discussed, along with methods for analyzing the sensor outputs. Finally, basic concepts on laser safety are presented. The range of processing parameters associated with each process are outlined. The impact of the basic laser characteristics such as wavelength, divergence, monochromaticity, coherence, polarization, intensity, stability, focusing, and depth of focus, as discussed in Part I, on each process, is emphasized.

The material in this book is suitable for a two-course sequence on laser processing. The material in Part I is adequate for an upper division/first-year graduate course in engineering. Parts II and III can then be used for a follow-up course, or the material in Part I can be skipped if only one course needs to be offered. In either case, Part II can be quickly reviewed and more time spent on Part III.

Two sets of nomenclature are used in this text. There is an overall nomenclature that is reserved for variables that are used throughout the text. In addition, each chapter has its own nomenclature that is used for variables that are used primarily in that chapter.

The author wishes to express his gratitude to all his colleagues and friends who have provided feedback on the manuscript. Special gratitude goes to all the graduate students who critiqued the course pack on which the book is based and to Mr. Rodney Hill (rodhillgraphics.com) for the skillful illustrations.

University of Michigan
Ann Arbor

Elijah Kannatey-Asibu, Jr.

ABOUT THE COMPANION WEBSITE

This book is accompanied by a companion website.

www.wiley.com/go/Kannatey-Asibu/PrinicplesLaserMaterialsProcessing

This website includes Solutions Manual

PART I
Principles of Industrial Lasers

1 Laser Background

1.1 LASER GENERATION

The term laser is an acronym for light amplification by stimulated emission of radiation, and thus a laser beam is a form of electromagnetic radiation. Light may be simply defined as electromagnetic radiation that is visible to the human eye. It has a wavelength range of about 0.37–0.75 μm, between ultraviolet and infrared radiation. Lasers, on the other hand, may have wavelengths ranging from 0.2 to 500 μm, that is, from X-ray to infrared radiation. In its simplest form, laser generation is the result of energy emission associated with the transition of an electron from a higher to a lower energy level or orbit within an atom.

Figure 1.1 illustrates the electromagnetic spectrum. The colors associated with the various wavelengths in the visible range are shown in Table 1.1.

1.1.1 Atomic Transitions

Under the right conditions, electrons within an atom can change their orbits. Light or energy (photon) is emitted as an electron moves from a higher level or outer orbit to a lower level or inner orbit and is absorbed when the reverse transition takes place. There is a specific quantum of energy (photon), ΔE, of specific wavelength or frequency associated with each transition from one energy level to another and is given by:

$$\Delta E = \frac{h_p c}{\lambda} = h_p \nu \tag{1.1}$$

where c is the velocity of light $= 3 \times 10^8$ (exactly 299,792,458) m/s, h_p is Planck's constant $= 6.625 \times 10^{-34}$ J s, λ is the wavelength (m), ν is the frequency of transition between the energy levels (Hz), and ΔE is the energy difference between the levels of interest.

1.1.1.1 Population Distribution For simplicity, let us focus our initial discussion on a single frequency, which corresponds to two specific energy levels, E_1 and E_2, where E_1 is the lower energy level and E_2 is a higher energy level, that is, $E_2 > E_1$. Furthermore, we let the *population* or number of atoms (or molecules or ions) per unit volume at level 1 be N_1, and that at level 2 be N_2. We also assume conditions of non-degeneracy. Degeneracy exists when there is more than one level with the same energy.

Principles of Laser Materials Processing: Developments and Applications, Second Edition. Elijah Kannatey-Asibu, Jr.
© 2023 John Wiley & Sons, Inc. Published 2023 by John Wiley & Sons, Inc.
Companion Website: www.wiley.com/go/Kannatey-Asibu/PrinicplesLaserMaterialsProcessing

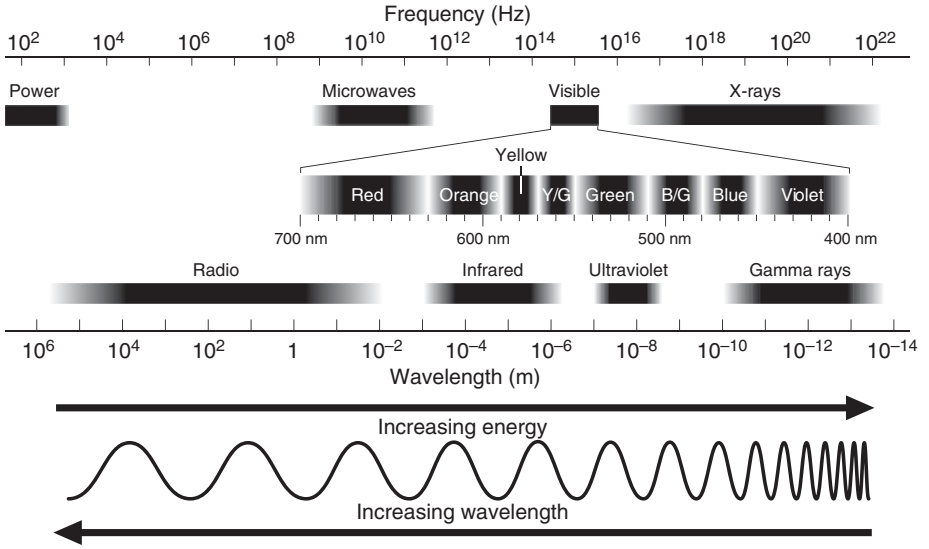


Figure 1.1 The electromagnetic spectrum.

Table 1.1 Wavelengths associated with the visible spectrum.

Wavelength range (nm)	Color
400–450	Violet
450–480	Blue
480–510	Blue-Green
510–550	Green
550–570	Yellow-Green
570–590	Yellow
590–630	Orange
630–700	Red

Under conditions of thermal equilibrium, the lower energy levels are more highly populated than the higher levels, and the distribution is given by Boltzmann's law that relates N_1 and N_2 as:

$$\frac{N_2}{N_1} = e^{-\frac{E_2-E_1}{k_B T}} = e^{-\frac{h_p \nu}{k_B T}} \tag{1.2}$$

where k_B is Boltzmann's constant $= 1.38 \times 10^{-23}$ J/K and T is the absolute temperature (K).

This is illustrated in Figure 1.2a. Figure 1.2b illustrates the equilibrium distribution for the more general case. Boltzmann's law holds for thermal equilibrium conditions.

1.1.1.2 Absorption When atoms in a ground state, E_1, are *excited* or *stimulated*, that is, subjected to some external radiation or photon whose energy is the same as the energy difference between E_1 and a higher state, E_2, the atoms will change their energy level, as shown in Figure 1.3a. This process is called *stimulated absorption*. The rate at which energy is absorbed by the atoms is given by:

$$n_{abs} = B_{12} N_1 e(\nu) \tag{1.3}$$

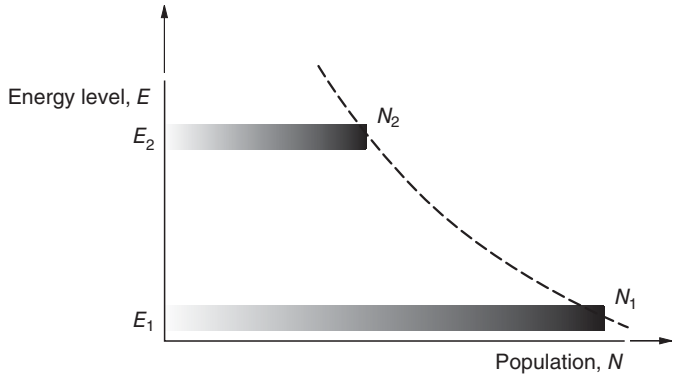


Figure 1.2 Schematic of Boltzmann's law. (a) Two-level system. (b) More general case for a multilevel system.

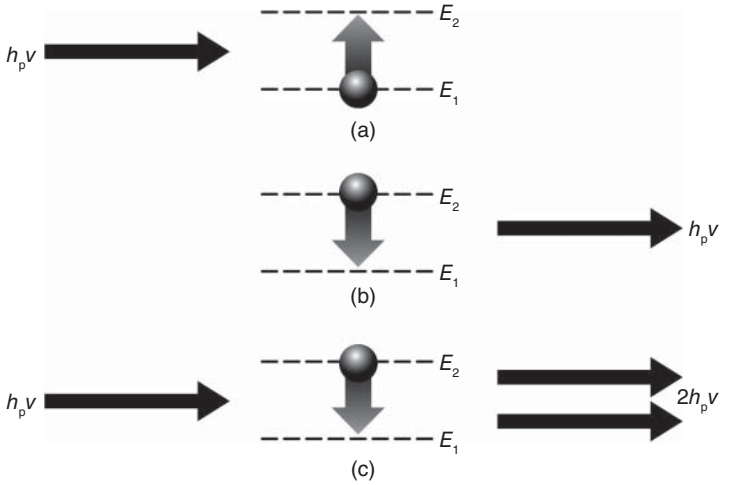


Figure 1.3 Schematic of (a) absorption, (b) spontaneous emission, and (c) stimulated emission.

where B_{12} is a proportionality constant referred to as the Einstein coefficient for stimulated absorption, or stimulated absorption probability per unit time per unit spectral energy density (m^3 Hz/J s), N_1 is the population of level 1 (per m^3), $e(\nu)$ is the energy density (energy per unit volume) at the frequency ν(J/m^3 Hz), and n_{abs} is the absorption rate (number of absorptions per unit volume per unit time).

Once the atom has been excited to a higher energy level, it can make a subsequent transition to a lower energy level, accompanied by the emission of electromagnetic radiation. The emission process can occur by *spontaneous emission* or *stimulated emission*. Each absorption removes a photon, and each emission creates a photon.

1.1.1.3 Spontaneous Emission Spontaneous emission occurs when transition from the excited to the lower energy level is not stimulated by any incident radiation (Figure 1.3b). The transition results in the emission of a photon of energy:

$$\Delta E = E_2 - E_1 = h_p \nu \tag{1.4}$$

where ν is the frequency of the emitted photon. In spontaneous emission, the rate of emission per unit volume, n_{sp}, to the lower energy level is only proportional to the population, N_2, at the higher energy level, and is independent of radiation energy density:

$$n_{sp} = A_e N_2 \tag{1.5}$$

where A_e is the Einstein coefficient for spontaneous emission, or spontaneous emission probability per unit time.

1.1.1.4 Stimulated Emission If the atom in energy level 2 is subjected to electromagnetic radiation or photon of frequency ν corresponding to the energy difference $\Delta E = E_2 - E_1$ between levels 1 and 2, the photon will stimulate the atom to undergo a transition to the lower energy level. The energy emitted is the same as the stimulating photon and is superimposed on the incident photon, thereby reinforcing the emitted light (Figure 1.3c). This results in stimulated emission, where the incident and emitted photons have the same characteristics and are in phase, resulting in a high degree of coherence, and the direction, frequency, and state of polarization of the emitted photon are essentially the same as those of the incident photon. The two photons can generate yet another set, with a resulting avalanche of photons. This is illustrated schematically in Figure 1.4. The rate of emission per unit volume, n_{st}, in the case of stimulated emission is given by:

$$n_{st} = B_{21} N_2 e(\nu) \tag{1.6}$$

where B_{21} is the Einstein coefficient for stimulated emission, or stimulated emission probability per unit time per unit energy density (m^3 Hz/J s).

1.1.1.5 Einstein Coefficients: A_e, B_{12}, B_{21} Under conditions of thermal equilibrium, the rates of upward ($E_1 \rightarrow E_2$) and downward ($E_2 \rightarrow E_1$) transitions must be the same. Thus, we have

$$n_{1\rightarrow 2} = n_{2\rightarrow 1} \tag{1.7}$$

or from Eqs. (1.3), (1.5), and (1.6),

$$B_{12} N_1 e(\nu) = B_{21} N_2 e(\nu) + A_e N_2 \tag{1.8}$$

This gives the energy density as:

$$e(\nu) = \frac{A_e}{B_{12}\frac{N_1}{N_2} - B_{21}} = \frac{A_e}{B_{12}\frac{h_p \nu}{k_B T} - B_{21}} \tag{1.9}$$

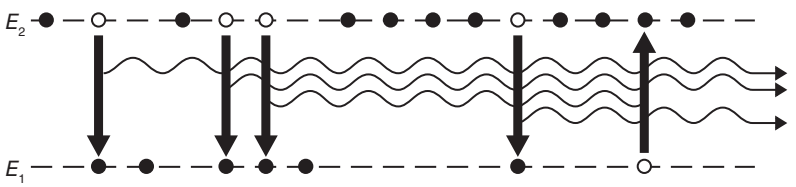


Figure 1.4 Illustration of the process of stimulated emission. Source: From Chryssolouris (1991)/Springer Nature.

This can be compared with the energy density from Planck's law on blackbody radiation:

$$e(\nu) = \frac{8\pi h_p \nu^3}{c^3} \frac{1}{\frac{h_p \nu}{e^{k_B T}} - 1} \tag{1.10}$$

Equations (1.9) and (1.10) are equivalent only if

$$B_{12} = B_{21} = B \tag{1.11a}$$

and

$$A_e = B\frac{8\pi h_p \nu^3}{c^3} \tag{1.11b}$$

Example 1.1 (a) Compare the rates of spontaneous and stimulated emission at room temperature ($T = 300$ K) for an atomic transition where the frequency associated with the transition is 3×10^{10} Hz, in the microwave region.

Solution:
From Eqs. (1.10) and (1.11),

$$\frac{h_p \nu}{k_B T} = \frac{6.625 \times 10^{-34}(\text{J s}) \times 3 \times 10^{10}(\text{Hz})}{1.38 \times 10^{-23}(\text{J/K}) \times 300(\text{K})} \approx 5 \times 10^{-3}$$

Thus,

$$\frac{A_e}{B} = (e^{5\times 10^{-3}} - 1)e(\nu) \approx 0$$

This indicates that the stimulated emission rate is much greater than the spontaneous emission rate, and thus amplification is feasible in the microwave range at room temperature.

(b) Repeat Example 1.1a for a transition frequency in the optical region of $\nu = 10^{15}$.

Solution:

$$\frac{h_p \nu}{k_B T} = \frac{6.625 \times 10^{-34}(\text{J s}) \times 10^{15}(\text{Hz})}{1.38 \times 10^{-23}(\text{J/K}) \times 300(\text{K})} \approx 160$$

$$\Rightarrow \frac{A_e}{B} = (e^{160} - 1)e(\nu) \approx \infty$$

indicating that spontaneous emission is then predominant, resulting in incoherent emission from normal light sources. In other words, under conditions of thermal equilibrium, stimulated emission in the optical range is very unlikely.

(c) What will be the wavelength of the line spectrum resulting from the transition of an electron from an energy level of 40×10^{-20} J to a level of 15×10^{-20} J?

Solution:
From Eq. (1.1), we have

$$\Delta E = \frac{h_p c}{\lambda} = h_p \nu$$

$$\Rightarrow (40 - 15) \times 10^{-20} = \frac{6.625 \times 10^{-34} \times 3 \times 10^8}{\lambda}$$

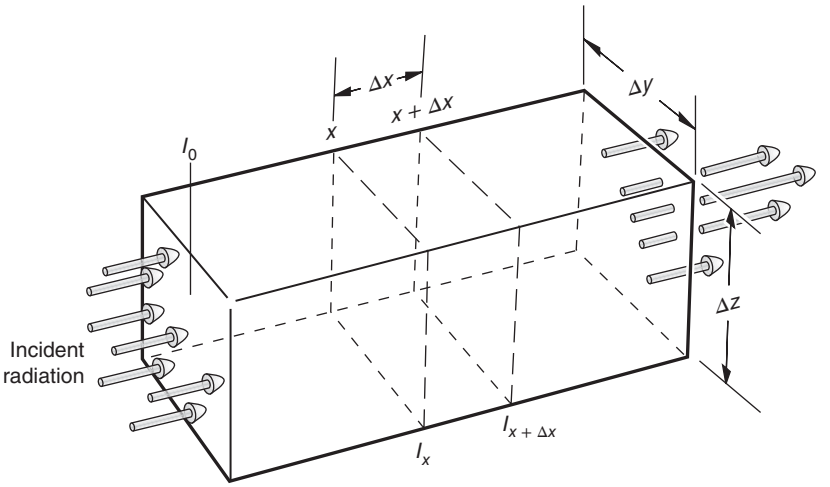


Figure 1.5 Propagation of a monochromatic beam in the x-direction.

Therefore,

$$\lambda = 0.792 \times 10^{-6}\ \text{m} = 0.792\ \mu\text{m}.$$

1.1.2 Lifetime

The *lifetime*, τ_{sp}, of atoms in an excited state is a measure of the time period over which spontaneous transition occurs. Strictly speaking, this is how long it takes for the number of atoms in the excited state to reduce to $1/e$ of the initial value. It can be shown that τ_{sp} is related to the Einstein coefficient for spontaneous emission by:

$$A_e = \frac{1}{\tau_{sp}} \tag{1.12}$$

1.1.3 Optical Absorption

The intensity, I, of a laser beam diminishes as it propagates through an absorbing medium. The variation of the beam intensity with distance (Figure 1.5) can be expressed as:

$$\frac{dI(x)}{dx} = -\alpha I(x) \tag{1.13}$$

Integration of Eq. (1.13) results in the following expression for the beam variation in the material:

$$I(x) = I_0 e^{-\alpha x} \tag{1.14}$$

where I_0 is the intensity of the incident beam (W/m^2) and α is the absorption coefficient (m^{-1}). This is known as the Beer–Lambert law and indicates that the beam intensity varies exponentially as it propagates into the medium.

Example 1.2 A medium absorbs 1% of the light incident on it over a distance of 1.5 mm into the medium. Determine

(i) The medium's absorption coefficient.
(ii) The length of the medium if it transmits 75% of the light.

Solution:

(i) From Eq. (1.14), we have

$$I(x) = I_0 e^{-\alpha x}$$

If 1% of the incident light is absorbed over a distance of 1.5 mm, then

$$0.99 \times I_0 = I_0 e^{-\alpha \times 1.5}$$

$$\Rightarrow \log_e 0.99 = -1.5 \times \alpha$$

Therefore,

$$\alpha = 6.7 \times 10^{-3}/\text{mm}$$

(ii) If 75% of the incident light is transmitted, then

$$0.75 \times I_0 = I_0 e^{-6.7 \times 10^{-3} \times x}$$

$$\Rightarrow \log_e 0.75 = -6.7 \times 10^{-3} \times x$$

Therefore,

$$x = 42.9 \text{ mm}$$

1.1.4 Population Inversion

The absorption coefficient, α, is positive if $N_1 > N_2$. The beam intensity then decreases exponentially with distance into the material. Since $N_1 > N_2$ under normal thermal equilibrium conditions, the beam will be attenuated as it propagates through the medium.

However, if conditions are such that the number of atoms at the higher energy level is greater than those at the lower energy level, that is, $N_2 > N_1$, then α will be negative, in which case the beam intensity will increase exponentially as it propagates through the medium. Equation (1.14) can then be written as:

$$I(x) = I_0 e^{\beta x} \tag{1.15}$$

β is referred to as the *small-signal gain coefficient* and is positive. Such amplification of the original radiation results from stimulated emission that occurs when the right conditions exist.

The condition where $N_2 > N_1$ with more atoms existing at the higher energy level than at the lower energy level is referred to as *population inversion* (Figure 1.6). Boltzmann's equation indicates that population inversion does not occur under normal thermal equilibrium conditions. For it to be possible, atoms within the laser medium have to be excited or *pumped* to a nonequilibrium state. This is done with the fusion of a substantial quantity of energy to the medium using an external source. A material in which population inversion is induced is called an *active medium*.

Even though population inversion is a necessary condition for achieving laser action, it is not a sufficient condition, since a significant number of the excited atoms decay spontaneously to the lower energy level. Compensation for the loss due to such decay is accomplished by introducing positive feedback into the system to amplify the laser beam. This is done using *optical resonators*.

1.1.5 Threshold Gain

In an actual laser, the active medium is normally placed between two mirrors, and these together constitute the resonator (Figure 1.7a). Initially, spontaneous emission results in photons being

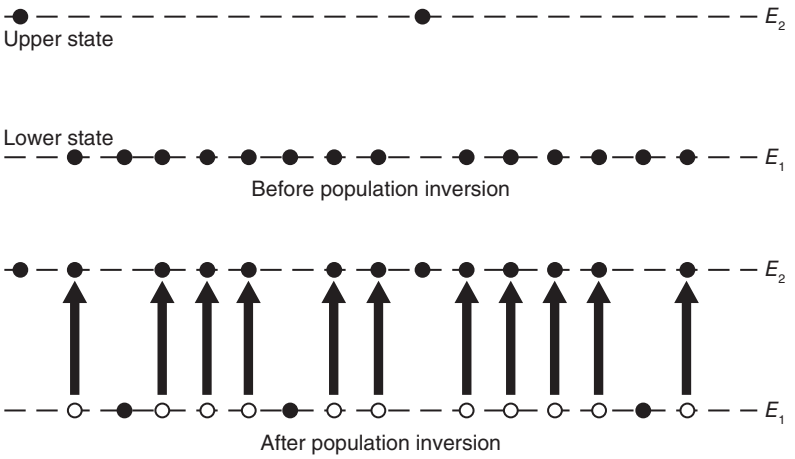


Figure 1.6 Population inversion. Source: From Chryssolouris (1991)/Springer Nature.

generated in all directions (Figure 1.7b). However, as stimulated emission becomes significant, an electromagnetic wave traveling along the axis of the resonator oscillates between the two mirrors (Figure 1.7c). When population inversion exists, the radiation is amplified on each passage through the medium resulting in buildup of signal intensity (Figure 1.7d). A useful output beam is obtained by making one of the mirrors partially transparent. This output beam is the laser that comes out of the system. Such output coupling, along with absorption and scattering by the active medium, absorption by the mirrors, and so on, result in losses in the radiation intensity. These losses may be simply described by Eq. (1.14). If the losses encountered by the radiation during each passage are greater than the amplification or *gain* of the laser, as shown in Eq. (1.15), then the oscillations cannot continue, and the radiation intensity will eventually die down. Thus, for oscillations to be maintained, the gain of the system must at least be equal to the losses in the system. This is accomplished when a *threshold gain*, β_{th}, is reached.

1.1.6 Two-Photon Absorption

When an electric field, E_1, is applied to a material, the polarization, P_1, that results is linear when the strength of the electric field is relatively low, as obtained for ordinary light sources, and is given by:

$$P_1(t) = \chi E_1(t) \tag{1.16}$$

where χ is the susceptibility of the dielectric medium and is independent of E_1(t), but it is a function of the frequency.

The actual relationship between the electric field and polarization is a power series given by:

$$P_1(t) = \chi^{(1)}E_1(t) + \chi^{(2)}E_1(t)^2 + \chi^{(3)}E_1(t)^3 + \cdots \tag{1.17}$$

Here, $\chi^{(\text{m})}$ is a tensor, and $\chi^{(1)} = \chi$, resulting in a linear term. However, $\chi^{(2)}$, $\chi^{(3)}$, and so on are nonlinear susceptibilities and define the degree of nonlinearity. The elements of $\chi^{(\text{m})}$ get smaller, the higher the m gets. Thus for the effects of the higher-order terms to be detected, the intensity of the light source has to be very high. This is normally achieved with pulsed lasers.

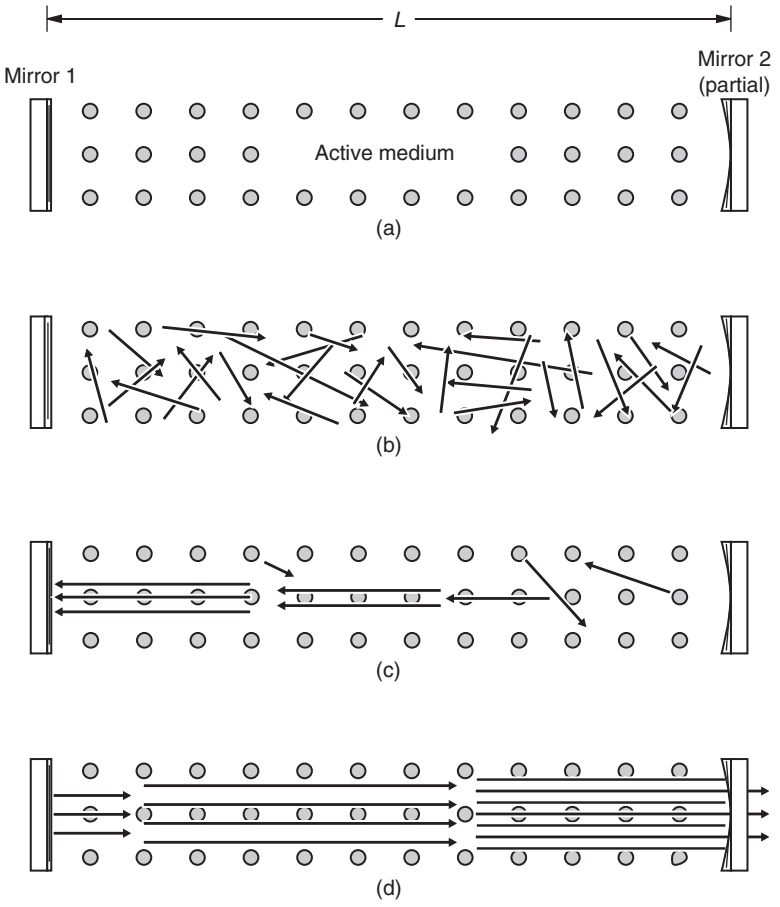


Figure 1.7 Illustration of laser amplification. Source: From Chryssolouris (1991)/Springer Nature.

Now normal excitation of an atom or molecule from a lower energy level to a higher level involves absorption of a photon of a specific energy or wavelength defined by the energy levels. When the atom is exposed to radiation of a high enough intensity, the atom can simultaneously absorb two longer wavelength photons, resulting in the same effect as a single photon of half the wavelength (Figure 1.8). The process is referred to as two-photon absorption. The combined energy of the two photons enables the atom to be excited to the higher energy level. In essence, each photon provides half the energy of the electronic transition. This occurs because each of the longer wavelength photons can excite the atom to a transient or virtual state which has a lifetime of a few femtoseconds. The virtual state is forbidden for a single-photon transition. Two-photon absorption can only take

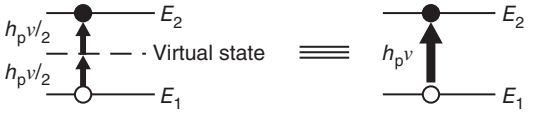


Figure 1.8 The two-photon absorption concept.

place if the second photon is absorbed before the virtual state decays. The concerted interaction of the two photons then results in a combined energy that enables the atom to be excited to a level equivalent to what would be induced by a single photon of half the wavelength or twice the energy.

Since the two photons have to be absorbed simultaneously, it is essential for the laser source to be ultrashort-pulsed, and of high peak power or intensity. Femtosecond lasers have thus been found very effective in a number of applications.

Two-photon absorption has found application in such areas as fluorescence microscopy, 3D optical data storage, lithography, photodynamic therapy, and microfabrication/nanofabrication.

1.2 OPTICAL RESONATORS

The principle of an optical resonator is illustrated schematically in Figure 1.7. To understand the characteristics of optical resonators, we take a brief look at wave propagation.

1.2.1 Standing Waves In A Rectangular Cavity

The electromagnetic waves of light have associated with them electric and magnetic fields which oscillate in a direction perpendicular to the general direction of wave propagation (Figure 1.9). The two waves are in phase, but perpendicular to each other, and can be represented by electric and magnetic vectors, $\mathbf{E_l}$ and $\mathbf{H}$, respectively.

Now let us consider a closed rectangular cavity with dimensions as shown in Figure 1.10, which is filled with a homogeneous and isotropic dielectric medium that sustains an electromagnetic wave. It can be shown that the permissible oscillating frequencies or waves of the field in the cavity are given by:

$$\nu^2 = \frac{c^2}{4}\left(\frac{m^2}{4a^2} + \frac{n^2}{4b^2} + \frac{p^2}{L^2}\right) \tag{1.18}$$

where m, n, and p give the number of nodes that the standing wave has along the y, z, and x axes, respectively. Each set of values of m, n, and p represents a well-defined cavity mode with a well-defined resonant frequency.

The number of modes, $M_c(\nu)$, that exist in a closed cavity of volume, V_c, over a frequency interval $\Delta\nu$ can be shown to be

$$M_c(\nu) = \frac{8\pi\nu^2 V_c}{c^3}\Delta\nu \tag{1.19}$$

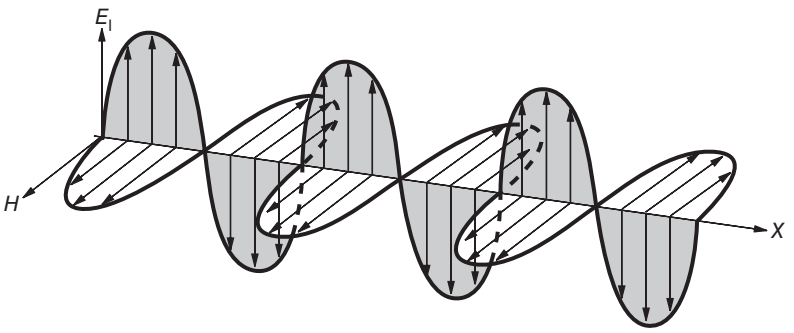


Figure 1.9 Electric and magnetic field vectors of an electromagnetic wave.

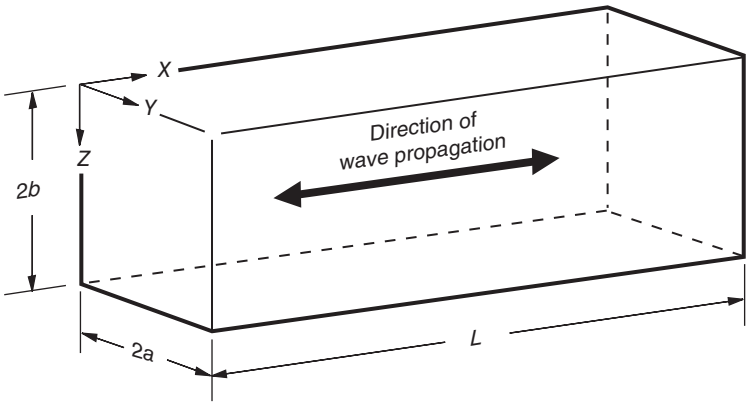


Figure 1.10 Rectangular cavity containing a dielectric medium.

Example 1.3 A laser cavity has a wavelength centered at $\lambda = 1500\,\text{nm}$. If the cavity volume is $2000\,\text{mm}^3$, determine the number of modes which fall in a bandwidth of $\Delta\lambda = 0.2\,\text{nm}$.

Solution:
The number of modes is given by:

$$M_c(\nu) = \frac{8\pi\nu^2 V_c}{c^3}\Delta\nu$$

Now

$$\nu\lambda = c, \text{ see Eq. (3.1a)}$$

where c is the velocity at which electromagnetic radiation propagates in free space.

$$\Rightarrow \nu = \frac{c}{\lambda} = \frac{3\times10^8}{1500\times10^{-9}} = 2\times10^{14}\,\text{Hz}$$

Also, from Eq. (3.2),

$$\Delta\nu = -\left(\frac{c}{\lambda^2}\right)\Delta\lambda = -\frac{3\times10^8}{(1500\times10^{-9})^2}\times0.2\times10^{-9} = -2.67\times10^{10}\,\text{Hz}$$

The negative sign arises since the wavelength increases as the frequency decreases. Now we have

$$M_c(\nu) = \frac{8\pi\times(2\times10^{14})^2\times2000\times10^{-9}}{(3\times10^8)^3}\times2.67\times10^{10} = 1.99\times10^9$$

1.2.2 Planar Resonators

An enclosed cavity with this high number of modes oscillating would result in a laser beam covering a very wide range of wavelengths. To overcome this problem, an open resonator (Figure 1.11), which has the lateral portions of the resonator removed, is normally used. With such a configuration, the modes traveling along the resonator axis will keep oscillating between the mirrors. However, the modes formed by superposition of waves traveling at an angle to the axis of the resonator will be lost after a few traversals. Thus, $m, n \ll p$. This is essentially a planar resonator.

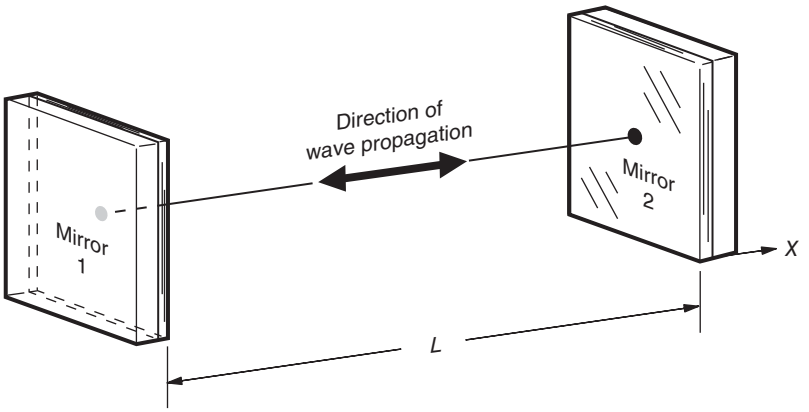


Figure 1.11 Planar or Fabry–Perot resonator.

We can then use a power series expansion to reduce Eq. (1.18), for the case where $a = b$, to give the resonant frequencies ν in the cavity, in the form:

$$\nu_{\text{mnp}} = \frac{\omega}{2\pi} \approx \frac{c}{2}\left[\frac{p}{L} + \left(\frac{m^2 + n^2}{a^2}\right)\frac{L}{8p}\right] \tag{1.20}$$

Ignoring the last two terms of Eq. (1.20) enables the resonant frequencies to be approximated by:

$$\nu = \frac{1}{2}\frac{pc}{L} \tag{1.21}$$

But the wavenumber k_x is given by:

$$k_x = \frac{2\pi}{\lambda} = \frac{p\pi}{L} \tag{1.22}$$

giving the oscillating wavelength as:

$$\lambda = \frac{2L}{p} \tag{1.23}$$

Thus, for the resonant or permissible frequencies that oscillate in the cavity, L is an integer multiple (p) of half-wavelengths.

The Fabry–Perot resonator is very difficult to operate in single mode and is not widely used.

1.2.2.1 Beam Modes There are two main categories of beam mode:

1. Longitudinal or axial modes.
2. Transverse modes.

Longitudinal Modes From Eq. (1.20), consider two modes that have the same values of m and n, but with a unit difference in the p value. The corresponding difference in frequency of oscillation is then

$$\Delta\nu_{\text{p}} = \frac{c}{2L} \tag{1.24}$$

These two contiguous modes will only differ in their field distribution along the longitudinal (or x) axis. Thus, the different modes resulting from the different values of p are referred to as the *longitudinal* or *axial modes*. The various axial modes appear as a single spot in the laser output, even though they are separated at discrete frequencies in the spectral domain.

For illustration, we note that a typical value for the mirror spacing in a laser would be about $L = 1500\,\text{mm}$. Then,

$$\Delta\nu_p \approx \frac{3 \times 10^{11}}{2 \times 1500} = 100\,\text{MHz}$$

Thus, the separation between the longitudinal modes for this cavity is 100 MHz (Figure 1.12a). The laser transition linewidth (Section 1.5.1; Figure 1.12b) is normally much wider, say, 1 GHz for a Doppler broadened medium (Section 1.5.2.2). As a result, the output of the laser will consist of a number oscillating discrete frequencies, the axial modes, within the broadened linewidth (Figure 1.12c).

Transverse Modes When the modes have the same p value but differ in their values for m and n, the resulting field distributions differ in the transverse direction and are thus referred to as the *transverse modes* of the cavity. These give an indication of the distribution of intensity within the beam cross section. If we consider adjacent transverse modes where $m = n$ but the value of p does not change, and with $\Delta m = 1$, then we have from Eq. (1.20) (Problems 1.12 and 1.13):

$$\Delta\nu_m = \frac{cL}{8a^2p}\left(m + \frac{1}{2}\right) \approx \upsilon_p \frac{\lambda L}{8a^2}\left(m + \frac{1}{2}\right) \tag{1.25}$$

And considering the same example used for the axial modes, but with the mirror dimensions $a = b = 10\,\text{mm}$, and with the beam in the infrared region of wavelength $\lambda \approx \frac{2}{3} \times 10^{-3}$ mm, then since $m \approx 1$, we find that

$$\Delta\nu_m \approx 100 \times \frac{1500 \times \frac{2}{3} \times 10^{-3}}{8 \times 100}\left(1 + \frac{1}{2}\right) \approx 0.2\,\text{MHz}$$

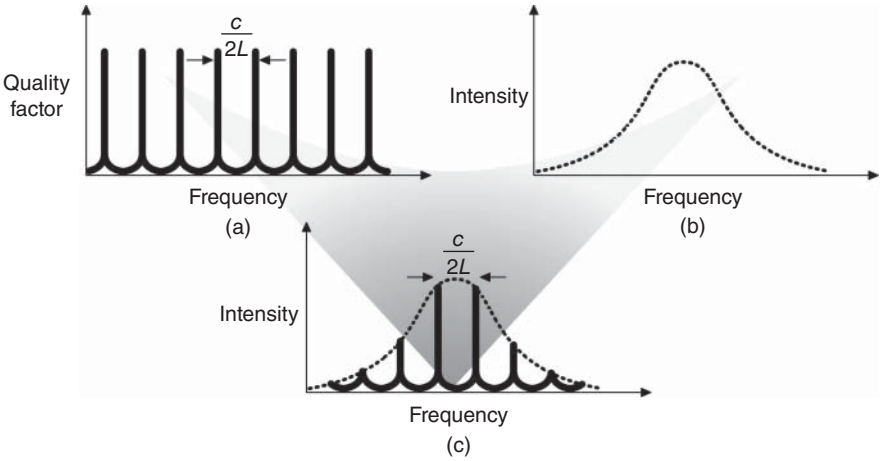


Figure 1.12 Longitudinal modes in a laser output. (a) Modes that can exist within the cavity. (b) Overall laser transition linewidth. (c) Output of the laser, indicating the modes within it. Source: From O'Shea et al. (1977)/with permission of Pearson.

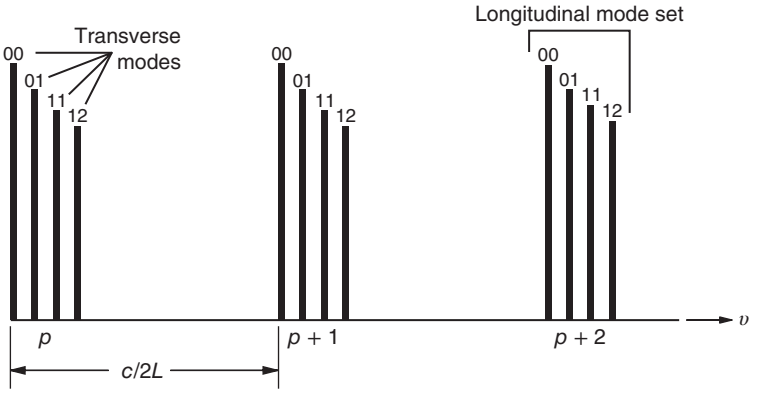


Figure 1.13 Resonant frequencies of a Fabry–Perot resonator.

Thus, $\Delta\nu_p \gg \Delta\nu_m$, indicating that the separation in frequency between longitudinal modes is much greater than that between transverse modes (Figure 1.13).

The transverse modes show as a pattern of spots and are often referred to as *transverse electromagnetic* or TEM modes. They are characterized by two integers (m, n) that indicate the number of modes in two orthogonal directions and are thus designated by TEM_{mn}. The m and n values are typically small, about 10 or less, and the larger the values, the more complex and spread out are the beam patterns.

The modes may have either rectangular or circular symmetry. For rectangular symmetry, the mode number is one less than the number of illuminated zones or pattern of spots in the corresponding direction. The first number corresponds to the y-axis, while the second number corresponds to the z-axis. Some common mode shapes are shown in Figure 1.14. TEM_{00} is the zero-order mode and ideally has a Gaussian distribution (Figure 1.15). A laser operating in this mode has the highest spectral purity and degree of coherence and can be focused to the laser's theoretical minimum radius.

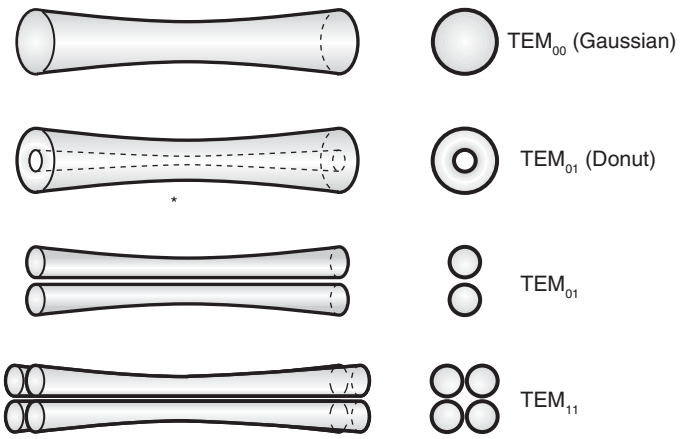


Figure 1.14 Mode shapes or schematic of TEM_{mn} mode patterns. Source: From Chryssolouris (1991)/Springer Nature.

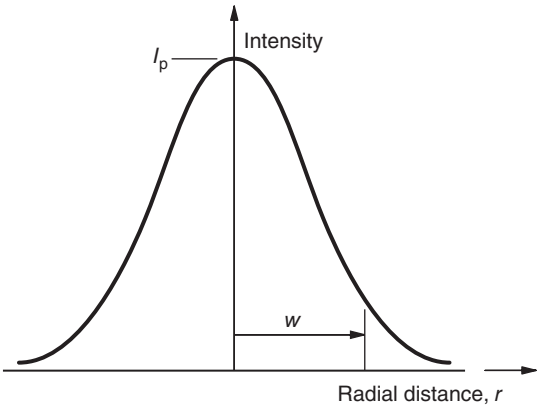


Figure 1.15 Gaussian (TEM_{00}) mode distribution.

This mode is more common in lower-power lasers. The distribution of laser intensity across the beam is then given by:

$$I(r) = I_p e^{\left(-\frac{2r^2}{w^2}\right)} \tag{1.26}$$

where $I(r)$ is the beam intensity at any radial position, r (W/cm^2), I_p is the peak beam intensity (W/cm^2), r is the radial coordinate (cm), and w is the radius at which beam intensity is $(1/e^2) \times I_p$ or $0.14 \times I_p$. That means 86% of the total energy is contained in a spot of radius w (cm).

In some manufacturing processes such as laser cutting, it is desirable to have the Gaussian mode so that the focused beam diameter which corresponds to the highest power density in the beam is achieved. On the other hand, the multimode beam is more desirable for laser butt welding applications where the wider beam area reduces restrictions on joint precision requirements.

1.2.2.2 Line Selection Lasers often undergo transition simultaneously on a number of transition lines or wavelengths. However, in situations where a higher degree of monochromaticity is desired, the laser should be made to oscillate on only one of the transitions. This can be done by inserting a wavelength dispersive element such as a prism or diffraction grating into the cavity (Figure 1.16). As the beam propagates through the prism, the individual wavelengths undergo different levels of refraction. The wavelength whose ray is directed normal to the mirror is reflected back into the

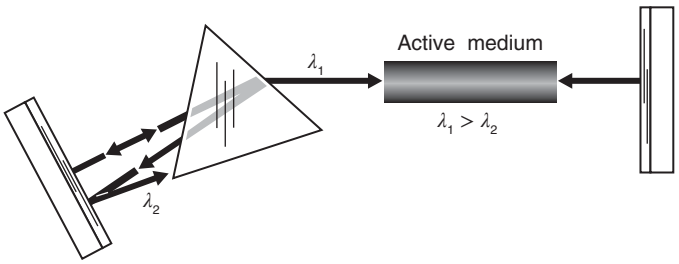


Figure 1.16 Wavelength selection.

cavity, while the other wavelengths experience losses, resulting in only one wavelength oscillating. Even when the laser is oscillating on a single line, that line itself may be broadened and could be oscillating on a number of modes.

1.2.2.3 Mode Selection When a laser is used for alignment, low beam divergence is necessary, and the lowest divergence is obtained with the TEM_{00} mode.

Transverse Mode Selection The higher-order modes can be eliminated by placing a diaphragm in the cavity and having it normal to the cavity axis. If the aperture of the diaphragm is made small enough, the higher-order modes are severely attenuated, leaving only the TEM_{00} mode. One drawback of this technique is reduction in the overall output power.

Longitudinal Mode Selection This is necessary for highly monochromatic output beams. One common technique involves cavity length variation.

As discussed earlier in Section 1.2.2.1, the longitudinal modes are spaced apart by:

$$\Delta\nu_{\mathrm{p}} = \frac{c}{2L} \tag{1.27}$$

Decreasing the cavity length L increases the spacing between the discrete frequency modes. Thus if the cavity length is decreased to the point that the frequency spacing is greater than the transition linewidth (Section 1.5.1; Figure 1.12), then only one mode will oscillate. This technique is effective for cavities where the laser linewidths are relatively small, such as for gas lasers. One setback of the technique, though, is that decreasing the cavity length decreases the volume of active material available for lasing, and thus the output power. It is not appropriate for solids and liquids where the laser linewidths are much greater.

Other methods for longitudinal mode selection include the Fabry–Perot etalon and Fox–Smith interferometer.

1.2.3 Confocal Resonators

A confocal resonator consists of a pair of spherical mirrors which have the same radius, r_{m}, and are positioned such that the focal points of the two mirrors coincide, that is, $F_1 \equiv F_2$. The center of curvature of each mirror then lies on the surface of the other mirror so that the distance L separating the mirrors is equal to the radius, r_{m} (Figure 1.17).

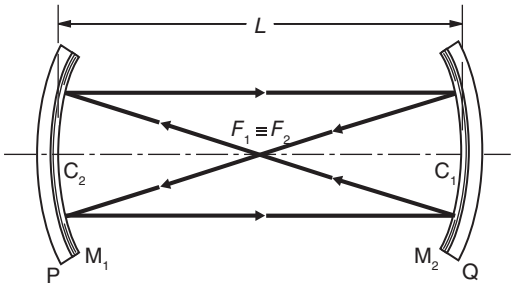


Figure 1.17 Confocal resonator. Source: From Svelto (2010)/Springer Nature.

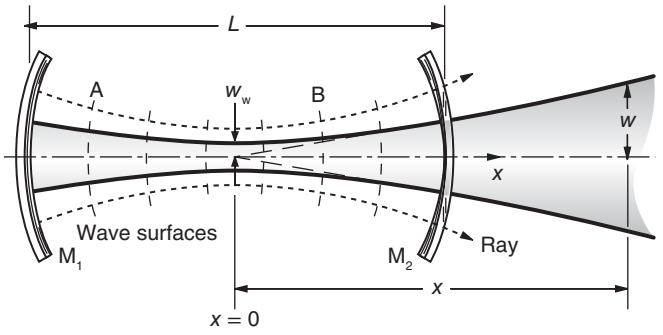


Figure 1.18 Beam size and phase for a TEM_{00} mode propagating in a confocal resonator.

The Gaussian field distribution for the confocal resonator is given by:

$$g_{00}(y,z) = e^{-\frac{\pi}{L\lambda}(y^2+z^2)}$$

The beam spot size on the mirror, w_m, is then given by:

$$w_m = \left(\frac{L\lambda}{\pi}\right)^{1/2} \tag{1.28}$$

and the spot size at the waist (radius of the beam cross section at the center of the cavity) is

$$w_w = \left(\frac{L\lambda}{2\pi}\right)^{1/2} \tag{1.29}$$

while that at any general location along the cavity axis, $w(x)$ is

$$w(x) = w_w\left[1 + \left(\frac{2x}{L}\right)^2\right]^{1/2} \tag{1.30}$$

This is illustrated schematically in Figure 1.18 by the solid curve.

The resonant frequencies in the confocal resonator cavity can be shown to be given by:

$$\nu_{mnp} = \frac{c(2p+m+n+1)}{4L} \tag{1.31}$$

Modes that have the same value of $(2p+m+n)$ have the same resonant frequency ν_{mnp} and are therefore frequency degenerate. The frequency difference between two contiguous longitudinal modes is given by:

$$\Delta\nu_p = \frac{c}{2L} \tag{1.32}$$

while that between contiguous transverse modes is given by:

$$\Delta\nu_m = \Delta\nu_n = \frac{c}{4L} \tag{1.33}$$

These are illustrated in Figure 1.19.

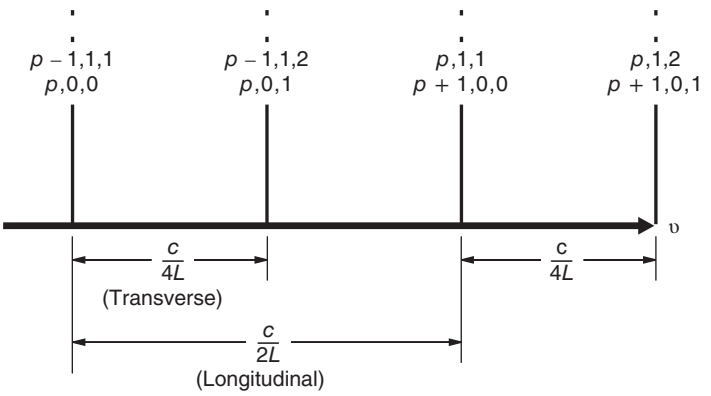


Figure 1.19 Oscillating frequencies of a confocal resonator. Source: From Svelto (2010)/Springer Nature.

The Rayleigh range (also known as the near-field distance), x_R, is defined as the distance between the beam waist and the point where the beam size is $\sqrt{2}$ times the waist size:

$$x_R \approx \frac{\pi w_w^{\,2}}{\lambda} \tag{1.34}$$

Distances within the Rayleigh range are referred to as the near-field region, while those much greater than the Rayleigh range are referred to as the far-field region.

One advantage of confocal resonators is that they exhibit the lowest diffraction losses for the TEM_{00} mode. One disadvantage, however, is that the *waist* spot size w_w that is obtained is normally much smaller than the cross-sectional area of the active medium so that the resonator does not make full use of the available medium or mode volume. As a result, confocal resonators are seldom used.

1.2.4 Concentric Resonators

With concentric resonators, the mirrors have the same radii of curvature, r_m, and are located such that the centers of curvature, C_1 and C_2, coincide (Figure 1.20). The separation distance, L, is then equal to twice the radius of curvature. Concentric resonators have the disadvantage that they are sensitive to mirror alignment. Furthermore, the beam spot size at the center of the resonator is small. As a result, they are not frequently used.

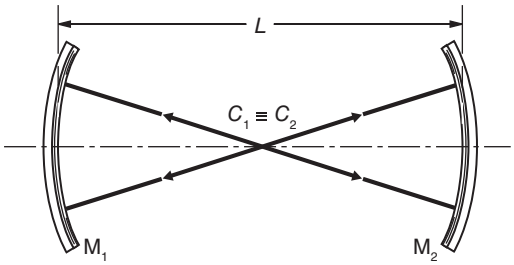


Figure 1.20 Concentric resonator. Source: From Svelto (2010)/Springer Nature.

Most resonators thus use concave mirrors with large radii of curvature, that is, $r_{mi} \gg L$. The spot sizes obtained for such mirrors are relatively large, and the resonators of which they are made are not as sensitive to alignment of the mirrors. Of course, as $r_{mi} \to \infty$, we end up with a plane mirror.

1.3 LASER PUMPING

Pumping is the process by which atoms are raised from a lower to a higher energy level to achieve a population inversion. Two techniques that are commonly used for industrial lasers are as follows:

1. Optical pumping.
2. Electrical pumping.

1.3.1 Optical Pumping

This form of pumping involves the excitation of the active medium using an intense source of light. Two types of light sources commonly used are as follows:

1. An arc lamp or a flashlamp.
2. Another laser, typically a diode laser.

1.3.1.1 Arc or Flashlamp Pumping Arc and flashlamps are preferably used with solid-state or liquid lasers since the lamp output bandwith is relatively broad, and thus for efficient absorption, the active medium also has to have a broad linewidth.

One form of lamp-based optical pumping is illustrated in Figure 1.21, where a linear pumping source (cylindrical in shape) is placed along one principal axis (focus, F_1) of an elliptical cavity (reflector) and the laser rod (active medium) along the other axis, F_2. When light from either axis (focus) of a specular (mirror-like) elliptical cavity in a plane orthogonal to the cylinder axis is reflected, it passes through the second axis. Thus, a significant portion of the light from the source is directed, after reflection, toward the laser rod. The reflector may be specular or diffuse. Diffuse reflectors tend to reflect light in all directions. Specular reflectors result in a higher pumping efficiency, but less uniform pumping of the active material, compared to diffuse reflectors.

Another type of configuration involves a lamp source in the form of a helix and is placed within a cylindrical container with a specular interior surface. The active material, which is in the form of a

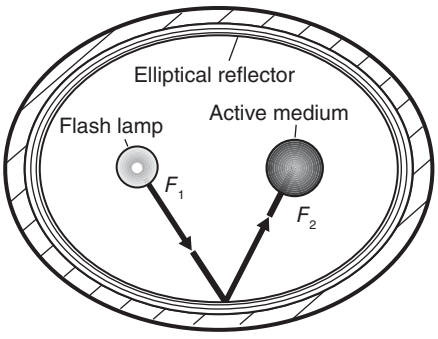


Figure 1.21 Schematic of an optical pumping system using an elliptical cavity.

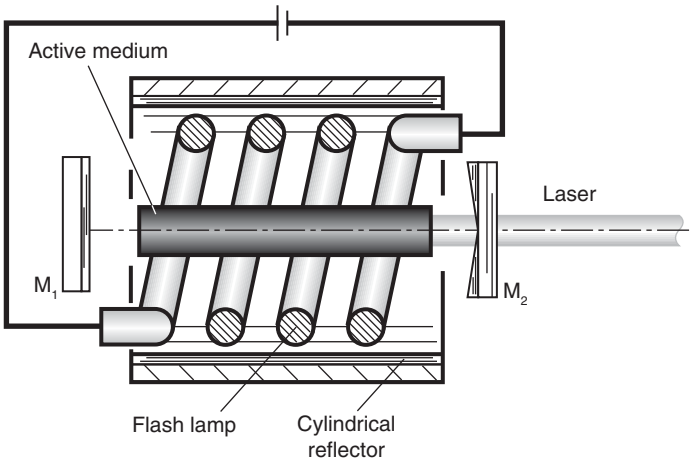


Figure 1.22 Schematic of an optical pumping system using a helical flashlamp.

rod, is placed at the center of the helix (Figure 1.22). Light from the lamp source propagates either directly to the rod, or after reflection from the interior of the cylindrical container.

Pulsed lasers are normally pumped using medium pressure (500–1500 Torr [mm Hg]) xenon or krypton lamps. An intense flash of light is applied for a period of the order of microseconds. Continuous-wave lasers are based on a high pressure (4000–8000 Torr) krypton or tungsten-iodine lamp that is continuously applied. Typical lamp life is about 800 hours.

1.3.1.2 Diode Laser Pumping Optical pumping efficiency is enhanced by using another laser for pumping. Since the pumping laser infuses energy at a specific wavelength, very little of its energy is wasted. Diode lasers are commonly used, especially for solid-state lasers. Two common pumping configurations used are as follows:

1. Longitudinal or end pumping.
2. Transverse or side pumping.

Longitudinal Pumping The longitudinal or end pumping technique is illustrated in Figure 1.23. The pumping beam is coupled to the laser medium from one end of the laser rod, along the resonator axis. A lens arrangement is used to concentrate the beam to a small and circular spot. The pumping beam can also be delivered using a fiber optic cable.

Transverse Pumping With transverse pumping, the laser diodes are radially arranged around the active medium (Figure 1.24.). Several diode pumped rods may be arranged in series in order to

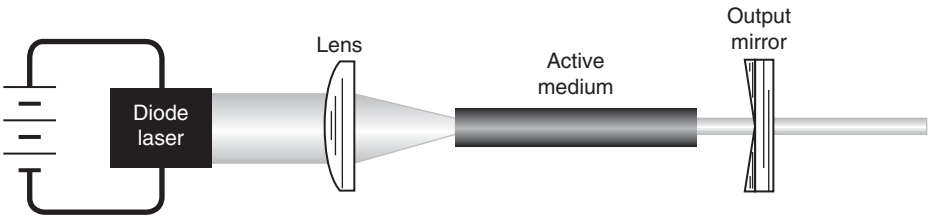


Figure 1.23 Longitudinal pumping using a diode laser.

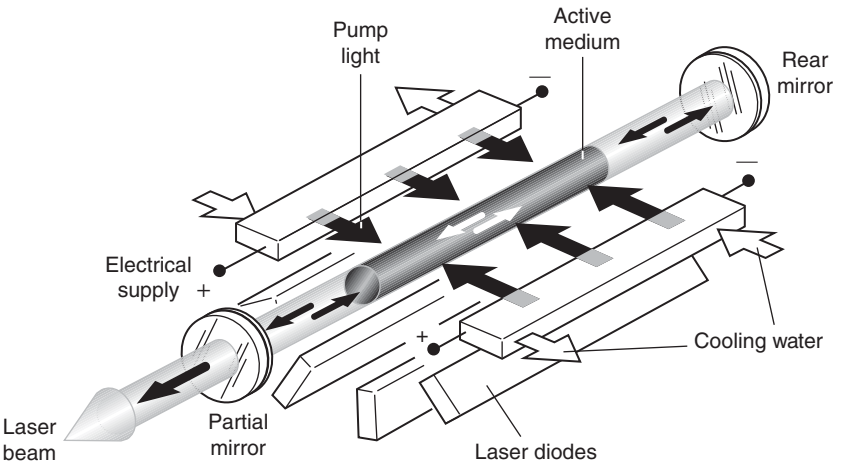


Figure 1.24 Transverse pumping using a diode laser. Source: ROFIN-SINAR Laser GmbH.

achieve high output powers in the kilowatt range. As in the case of longitudinal pumping, the beam can also be delivered using fiber optic cables.

The efficiency of a lamp-based system is normally in the range of 1–3%. This is the fraction of the input electrical energy that is converted to useful output. The 97–99% of energy that is not used is removed in the form of heat using a chiller. Using a diode laser for pumping increases the efficiency to the range of 30–40%.

1.3.2 Electrical Pumping

Electrical pumping is more commonly used with gas and semiconductor lasers. Pumping involves the creation of an electrical discharge by passing electrical energy, of the order of 1 to 2 kV at a low current of about 50 mA, through the gas medium, in the case of gas lasers (Figure 1.25).

Acceleration of electrons by the electric field results in excitation of atoms, ions, or molecules with which they collide, from the ground state to a higher level by imparting their kinetic energy to the atoms. The ground state refers to the situation when all the electrons are in their normal orbits.

The electrical energy may be applied as a dc (direct current) or rf (radio frequency alternating) electric field. In dc-generated discharges, the electrodes are contained within the discharge tube, and over a period of time, the electrodes may become contaminated. Changing the electrodes requires

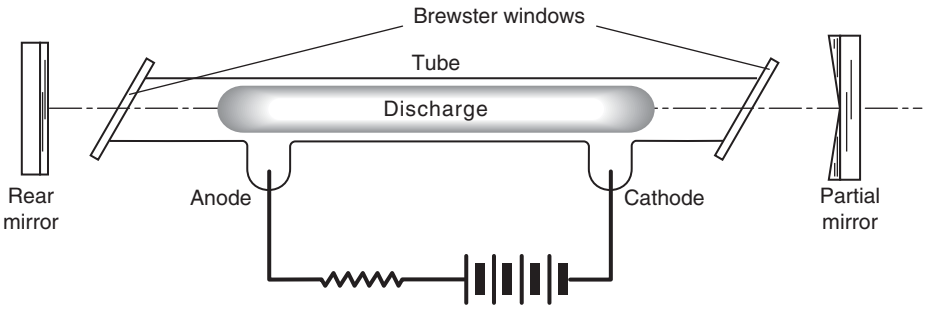


Figure 1.25 Schematic of an electrical pumping system.

access to the interior of the tube, a process that may in itself introduce additional contamination. On the other hand, with rf-generated discharges, the electric field is generated from outside the discharge tube, and thus the electrodes can be changed without accessing the interior of the tube.

Excitation frequencies for rf-excited lasers are typically of the order of 30 MHz, and thus they tend to have a fast response to a change in power, as compared to dc-excited lasers. This characteristic makes it easier to modulate rf-excited lasers, resulting in very high pulse rates, up to 15 kHz. In addition, higher levels of electrical power densities can be achieved, along with greater homogeneity of the plasma, in rf-excited lasers. This further improves beam quality and also enables more compact generator configuration to be achieved for the same power output. Radio frequency excitation does not require the use of permanent anodes and cathodes. Thus, cathode problems that result from gas reactions are eliminated. However, rf-excited lasers tend to have lower overall efficiencies compared to dc-excited lasers.

1.4 SYSTEM LEVELS

1.4.1 Two-Level System

As indicated in Eq. (1.2), the distribution of atoms under thermal equilibrium at any given temperature is given by Boltzmann's law which may be expressed in general form as:

$$\frac{N_j}{N_i} = e^{-\frac{\Delta E_{ij}}{k_B T}} \tag{1.35}$$

where N_i and N_j are the number of atoms per unit volume in the lower and higher energy states, respectively, ΔE_{ij} is the energy difference between the two states, k_B is Boltzmann's constant, and T is the absolute temperature.

It can be shown that for a two-level system, absorption occurs only insofar as there are more atoms in the lower energy state than in the higher energy state. As the atoms are pumped to the point where the two energy levels are equally populated, absorption stops. Thus, population inversion is difficult to achieve in a two-level system.

1.4.2 Three-Level System

We now consider a three-energy-level system (Figure 1.26), with populations per unit volume given by N_1, N_2, and N_3. The lowest energy level is 1, and 3 is the highest. Atoms are normally pumped from the lowest level, 1, to the highest level, 3. From there, they undergo rapid decay to level 2 and radiative transition that results in lasing action occurs between levels 2 and 1.

Let W_{ij} be the probability per unit time of the atom being thermally excited between levels i and j.

A population inversion exists between energy levels 1 and 2 for a three-level system, that is $N_2 > N_1$, only if

$$W_{32} > W_{21} \tag{1.36}$$

However, this is only a necessary condition for laser oscillations to occur, but not sufficient since the population inversion has to be high enough to overcome excess cavity losses. A threshold level of the population inversion must also be achieved. Now from Eq. (1.12), we have

$$W_{ij} \propto \frac{1}{\tau_{ij}} \tag{1.37}$$

where τ_{ij} is the lifetime for transition from level i to level j.

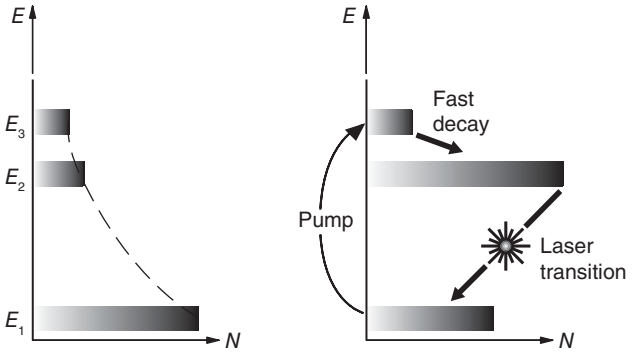


Figure 1.26 Schematic of a three-level system. (a) No pumping. (b) Intense pumping. Source: From O'Shea et al. (1977)/with permission of Pearson.

Thus, the lifetime for transition from level 3 to level 2 must be smaller than the lifetime for transition from level 2 to level 1, for population inversion to be possible in a three-level system (Figure 1.26). Level 3 is then practically empty, while population buildup occurs on level 2, resulting in population inversion between levels 2 and 1.

Lasing transition occurs between levels 2 and 1, and since level 3 is normally not a laser level, it can have a broadband, and that makes it possible to pump the system effectively using an optical source with a broad bandwidth.

1.4.2.1 Four-Level System For the four-level laser (Figure 1.27), we denote the levels of interest as 0, 1, 2, and 3, for the lasing levels to correspond with those of the three-level system. It can then be shown that population inversion between energy levels 1 and 2 can be achieved, that is $N_2 > N_1$, only if

$$W_{10} > W_{21} \tag{1.38}$$

Population inversion can be obtained in a four-level system much more easily, and with very little pumping power because the process is independent of pumping power, while that for a three-level system depends on the pumping power.

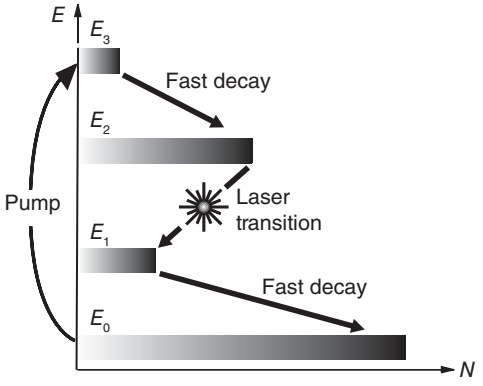


Figure 1.27 Schematic of a four-level system. Source: From O'Shea et al. (1977)/with permission of Pearson.

Just as for the three-level system, level 3 in the four-level system can have a broadband since it is not a laser level, and thus it is possible to pump the system effectively using an optical source with a broad bandwidth. Furthermore, the energy difference between levels 1 and 0 has to be much greater than k_BT, that is, $\Delta E_{01} \gg k_B$T so that the population of level 1 is essentially negligible at ordinary temperatures. Otherwise, level 1 would be populated by atoms from level 0 through thermal excitation, thereby reducing the population inversion and the efficiency of the system.

1.5 BROADENING MECHANISMS

Our discussion thus far has been based on the idealized assumption that there exists a single frequency, the resonant frequency, at which lasing occurs. However, the transition frequency is not exactly monochromatic, but it is spread over a finite frequency range. The principal mechanisms that cause broadening of the transition frequency and their corresponding line shape functions are discussed in Sections 1.5.1–1.5.3.

1.5.1 Line Shape Function

The finite range over which the transition frequency is spread is called the *atomic linewidth* and is centered about a resonant frequency, ω_0. The interaction between the atom and incident radiation is thus a function of frequency and is known as the *line shape function*, designated $g(\omega, \omega_0)$. The function $g(\omega, \omega_0)d\omega$ represents the probability that the transition frequency lies between ω and $\omega+d\omega$ and is centered at the resonant frequency ω_0. In the normalized form, the line shape function is defined such that

$$\int_{-\infty}^{\infty} g(\omega, \omega_0)d\omega = 1 \tag{1.39}$$

It is characterized using the width or frequency range $\Delta\omega_0 = 2\Delta\omega = 2(\omega - \omega_0)$ at the point of the curve where its amplitude is half the maximum amplitude. This is referred to as the full width at half maximum (FWHM) (Figure 1.28).

1.5.2 Line-Broadening Mechanisms

Broadening mechanisms can be grouped under two main categories:

1. Homogeneous broadening.
2. Inhomogeneous broadening.

1.5.2.1 Homogeneous Broadening In homogeneous broadening, the mechanism has the same effect on each atom, with each one having the same resonant frequency and atomic line shape. Homogeneous broadening mechanisms include natural or intrinsic broadening, and collision broadening.

Natural Broadening Natural broadening is the basic form of broadening that would exist even if there were no other form of broadening. It is associated with the finite lifetime of the excited state due to spontaneous emission, and the corresponding line shape function is given by:

$$g(\omega, \omega_0) = \frac{2\tau_{sp}}{\pi} \frac{1}{1 + 4\tau_{sp}{}^2(\omega - \omega_0)^2} \tag{1.40}$$

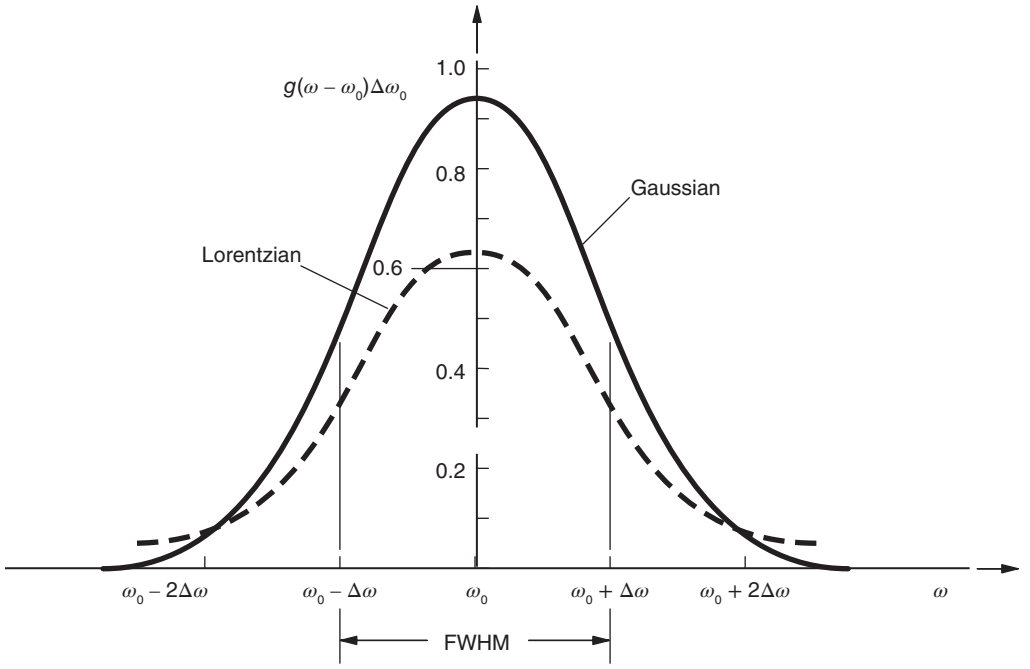


Figure 1.28 Lorentzian and Gaussian line shape functions.

This line shape function is Lorentzian (Figure 1.28). The maximum value occurs at $\omega = \omega_0$, and is $2\tau_{sp}/\pi$. The FWHM is

$$\Delta\omega_0 = 2\Delta\omega = \frac{1}{\tau_{sp}} \tag{1.41}$$

Natural broadening sets a lower limit on the atomic linewidth, and because of its relatively small magnitude, it tends to be predominated by other broadening mechanisms.

Collision Broadening Collision broadening is more predominant in gaseous media and results primarily from collision between atoms. The phase ϕ of the radiated wave changes abruptly if the atom suffers a collision while undergoing emission (Figure 1.29). Such random phase changes cause broadening of the atomic linewidth.

ϕ is constant between collisions but changes randomly at each collision. The line shape function of the emission associated with such collisions is given by:

$$g(\omega, \omega_0) = \frac{1}{\pi} \frac{t_{ca}}{1 + (\omega - \omega_0)^2 {t_{ca}}^2} \tag{1.42}$$

It is also Lorentzian. The maximum value of the function at $\omega = \omega_0$ is t_{ca}/π. The corresponding FWHM is then given by:

$$\Delta\omega_0 = 2\Delta\omega = \frac{2}{t_{ca}} \tag{1.43}$$

An estimate of the average time between collisions of an atom or molecule, t_{ca}, can be obtained by considering the mean free path and the average thermal velocity and is given by:

$$t_{ca} = \frac{\text{Mean free path}}{\text{Average thermal velocity}} = \frac{(m_m k_B T)^{\frac{1}{2}}}{16\pi^{\frac{1}{2}} P_g {r_a}^2} \tag{1.44}$$

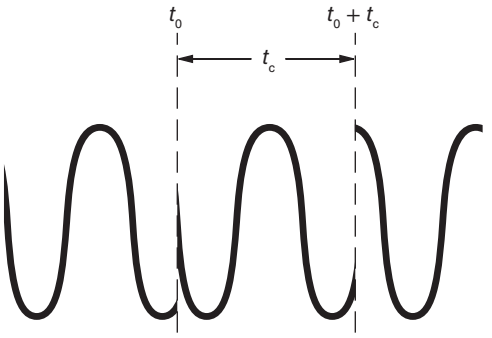


Figure 1.29 Phase changes associated with radiation emission resulting from atomic collision.

where k_B is Boltzmann's constant, m_m is the atomic or molecular mass, P_g is the gas pressure, r_a is the radius of the atom or molecule, and T is absolute temperature.

For example, for neon gas which is at room temperature and under a pressure of 0.5 Torr, we have

$$t_{ca} \approx 0.5 \times 10^{-6}\text{s}$$

The corresponding FWHM is then

$$\Delta\omega_0 = 2\Delta\omega = \frac{2}{0.5 \times 10^{-6}} = 4 \times 10^6 \text{ rad/s}$$

or

$$\Delta\nu_0 = 0.64 \text{ MHz}$$

Similar effects arise in solids as a result of the interaction of atoms with the lattice.

1.5.2.2 Inhomogeneous Broadening In inhomogeneous broadening, the resonant frequency of each atom is different or spread out, resulting in a band of frequencies for the entire ensemble, even though the lines of individual atoms are not necessarily broadened.

The most common form of inhomogeneous broadening in gas lasers is Doppler broadening. It is essentially a result of the *Doppler effect*, which is the phenomenon whereby an observer who is moving at a velocity **u** detects a frequency ω_0 (centerline frequency) that is different from the frequency ω detected by an observer who is in a stationary frame. A simple illustration of this is the sound of a moving car as detected by an observer standing by the roadside. Even though the frequency of the sound is the same to the driver (observer in the car, ω_0), the sound heard or detected by the standing observer (ω) has a higher frequency as the car approaches and a lower frequency as the car recedes from the stationary observer. This is because the number of oscillations that reach the stationary observer per unit time is increased or decreased depending on whether the car is approaching or receding.

Gaseous atoms, as we know, are in constant motion at rather high velocities. Now let the radiation emitted by an atom while in motion have a frequency ω_0 (as measured in the moving coordinate system or as seen by the atom). Further, let the absolute velocity of the atom (as measured relative to the stationary coordinate system) be u. If the component of this velocity in the direction of

wave propagation is u_x, then the frequency ω of the emitted radiation as observed in the stationary coordinate system will be related to ω_0 by:

$$\omega = \omega_0 \left(1 \pm \frac{u_x}{c}\right) \tag{1.45}$$

The positive or negative sign depends on whether the velocity is in the opposite or same direction as the emitted wave. Thus, even though the transition or centerline frequency ω_0 might be constant, each atom would be observed to be emitting radiation at a different resonant frequency depending on its direction, and that is what results in broadening of the emission line shape.

The line shape function for Doppler broadening has a Gaussian distribution and is given by:

$$g(\omega, \omega_0) = \frac{c}{\omega_0}\left(\frac{m_\mathrm{m}}{2\pi k_\mathrm{B} T}\right)^{\frac{1}{2}} e^{\left[-\frac{m_\mathrm{m} c^2}{2k_\mathrm{B} T}\left(\frac{\omega-\omega_0}{\omega_0}\right)^2\right]} \tag{1.46}$$

It has a maximum at $\omega = \omega_0$ given by $\frac{c}{\omega_0}\left(\frac{m_\mathrm{m}}{2\pi k_\mathrm{B} T}\right)^{\frac{1}{2}}$, and the FWHM is given by:

$$\Delta\omega_0 = 2\Delta\omega = 2\omega_0\left(\frac{2k_\mathrm{B} T \ln 2}{m_\mathrm{m} c^2}\right)^{\frac{1}{2}} \tag{1.47}$$

1.5.3 Comparison of Individual Mechanisms

To get an idea of the relative impacts of each of these broadening mechanisms, we consider a He–Ne laser with a transition at $\lambda = 632.8$ nm, and which is at room temperature, 300 K and pressure of 0.5 Torr. For natural broadening, $\tau_\mathrm{sp} \approx 10^{-8}$ s. Thus,

$$\Delta\nu_{0\mathrm{nat}} = 2\Delta\nu_\mathrm{nat} = \frac{1}{2\pi\tau_\mathrm{sp}} \approx 20\ \mathrm{MHz}$$

For collision broadening, $t_\mathrm{ca} \approx 0.5 \times 10^{-6}$ s. Thus,

$$\Delta\nu_{0\mathrm{col}} = 2\Delta\nu_\mathrm{col} = \frac{1}{\pi t_\mathrm{ca}} \approx 0.64\ \mathrm{MHz}$$

For Doppler broadening,

$$\Delta\nu_{0\mathrm{Dop}} = 2\Delta\nu_\mathrm{Dop} = 2\nu_0\left(\frac{2k_\mathrm{B} T \ln 2}{m_\mathrm{m} c^2}\right)^{\frac{1}{2}} \approx 1.7\ \mathrm{GHz}$$

The example indicates that Doppler broadening is the predominant form, with collision broadening being relatively negligible. This is the case for the He–Ne laser where the cavity pressures are relatively low. However, at high enough pressures, such as with the CO_2 laser, collision broadening becomes significant.

1.6 BEAM MODIFICATION

Some of the basic characteristics of a pulsed beam, such as the peak power, pulse duration, and pulse repetition rate (pulse frequency) can be changed during processing to enable desired beam properties

to be achieved. For example, a higher peak power may be required to initiate cutting or welding in some materials, with the power level that is needed being significantly reduced once the material melts. Sections 1.6.1–1.6.3 discuss techniques used for varying the beam characteristics.

1.6.1 Quality Factor

The Q-factor or quality factor, Q_f, is a measure of the losses or energy dissipation in any mode of the laser cavity and is defined as:

$$Q_f = \frac{\omega_0 \times \text{energy stored in the mode}}{\text{Energy lost or dissipated in the mode per unit time}} \tag{1.48}$$

$$= \frac{2\pi \times \text{energy stored in the mode}}{\text{Energy lost or dissipated per cycle}} = \frac{\omega_0 \times Q}{q_l} \tag{1.49}$$

where Q is the energy stored in the mode at any instant in time, q_l is the power loss in the mode, and ω_0 is the oscillation or transition frequency of that mode.

A cavity whose Q-factor is high stores energy well. Equation (1.49) can be rewritten as:

$$q_l = \frac{\omega_0 \times Q}{Q_f} \tag{1.50}$$

or

$$\frac{dQ}{dt} = -\frac{\omega_0 \times Q}{Q_f} \tag{1.51}$$

The negative sign reflects a loss or reduction in energy with time. The solution of Eq. (1.51) is

$$Q(t) = Q_0 e^{-\frac{\omega_0}{Q_f}t} \tag{1.52}$$

where Q_0 is the initial energy at time $t = 0$.

The time constant of the system described by Eq. (1.52), which is equivalent to the cavity lifetime, τ_c, is thus given by:

$$\tau_c = \frac{Q_f}{\omega_0} \tag{1.53}$$

The corresponding electric field amplitude of the mode is given by:

$$E_l(t) = E_0 e^{-\frac{1}{2\tau_c}t} e^{i\omega_0 t} = E_0 e^{\left(i\omega_0 - \frac{\omega_0}{2Q_f}\right)t} \tag{1.54}$$

since the time constant of the field amplitude is twice that of the field intensity. E_0 in Eq. (1.54) is the initial electric field amplitude at time $t = 0$. The first exponential expresses the decay of the signal intensity with time.

Taking the Fourier transform gives the field spectral density as:

$$E_l(\omega) = \frac{1}{2\pi}\int_0^\infty E_l(t)e^{-i\omega t}dt = \frac{1}{2\pi}\frac{E_0}{\frac{\omega_0}{2Q_f} + i(\omega - \omega_0)} \tag{1.55}$$

The spectral density of the signal intensity then becomes

$$I(\omega) = |E_l(\omega)|^2 = \frac{1}{4\pi^2}\frac{{E_0}^2}{\left(\frac{\omega_0}{2Q_f}\right)^2 + (\omega - \omega_0)^2} \tag{1.56}$$

This is a Lorentzian distribution with a FWHM $\Delta\omega_0$ given by:

$$\Delta\omega_0 = \frac{\omega_0}{Q_\mathrm{f}} \tag{1.57}$$

The quality factor can be shown to be given as:

$$Q_f = \frac{2\omega_0 L}{c_m(2\alpha L - \ln R_1 R_2)} \tag{1.58}$$

where R_1 and R_2 are the reflection coefficients of the two mirrors, L is the length of the active medium, $c_\mathrm{m} = c/n$ is the velocity of light in the medium, and n is the medium refractive index.

1.6.2 Q-Switching

Some basic pulsed solid-state lasers have a pulse duration of about 1 ms, with peak powers of the order of kilowatts. Q-switching is a technique that is used to produce laser outputs of higher power (megawatts), and of shorter duration (nanoseconds). It must be noted, however, that even though the output power is increased, the total energy content of the pulse is not and may even be less.

To understand the principle behind Q-switching, we consider a laser cavity in which a shutter is placed in front of one of the mirrors. When the shutter is closed, it prevents light energy from reaching the second mirror, and thereby being reflected back into the cavity. In other words, most of the light energy is lost, and oscillation cannot take place. The Q-value of the cavity is then very low (Figure 1.30a-1), since the losses are high (Figure 1.30a-2). As pumping of the laser continues (Figure 1.30a-3), the population inversion keeps building up (Figure 1.30a-4), far in excess of the threshold value, without any oscillation taking place. When a significantly high value of population inversion has been achieved, the shutter is suddenly opened to reduce the losses. At this point, the gain of the laser (due to the high population inversion) is much greater than the losses. The high energy accumulated as a result of the large difference between the instantaneous and threshold population inversions is then released as an intense beam of short duration (Figure 1.30a-5). Opening of the shutter to reduce the losses increases the Q-value of the cavity. Hence, the name Q-switching.

Q-switching can either result in a single pulse, in which case the pump rate is also pulsed (Figure 1.30a) or it can be repetitively pulsed, in which case pumping is continuous (Figure 1.30b). The necessary conditions for a laser to be Q-switchable are as follows:

1. The lifetime, τ_u, of the upper level has to be longer than the cavity buildup time, t_cb:

$$\tau_\mathrm{u} > t_\mathrm{cb}$$

   This enables the upper level to store the extra energy pumped into it.
2. The pumping duration, t_p, has to be longer than the cavity buildup time, t_cb. Preferably, it has to be at least as long as the lifetime of the upper level, τ_u:

$$t_\mathrm{p} > t_\mathrm{cb} \text{ and } t_\mathrm{p} \geq \tau_\mathrm{u}$$

3. The initial cavity loss must be high enough during pumping to prevent oscillation during that period.
4. The change in Q_f value must be sudden.

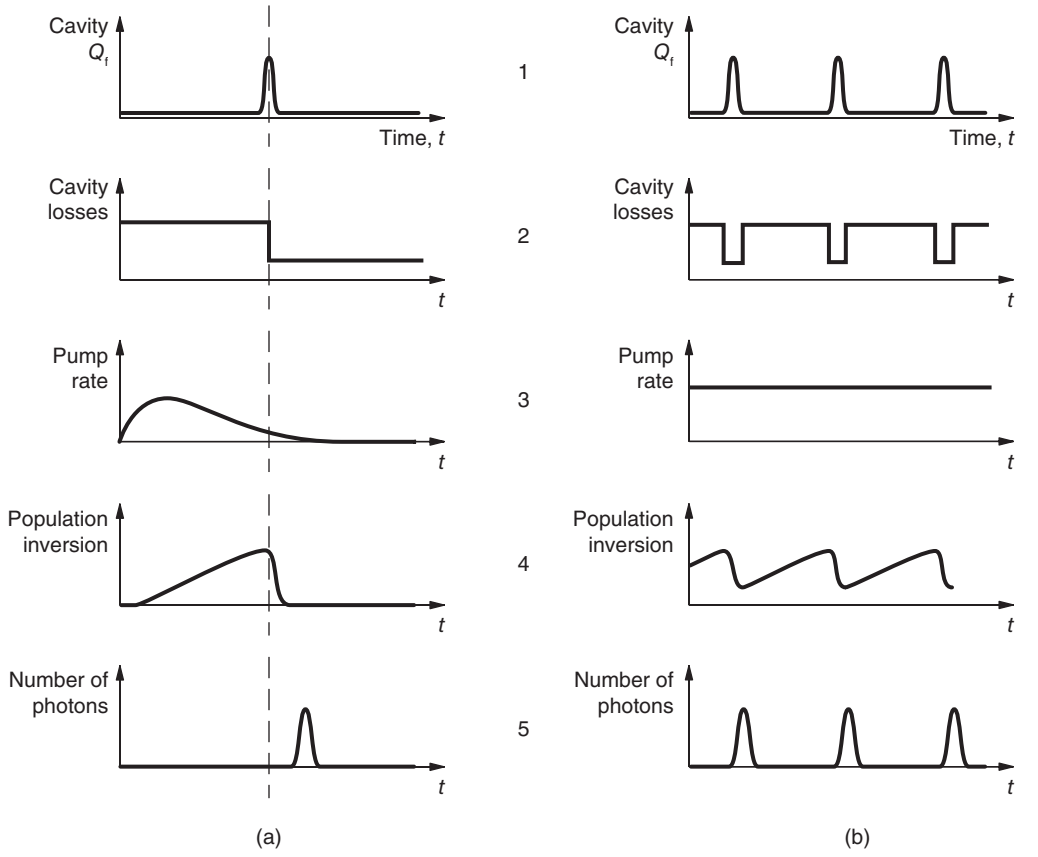


Figure 1.30 Time evolution of the Q-switching process: (a) single pulse and (b) continuous pulsing. Source: From Svelto (2010)/Springer Nature.

The major techniques for Q-switching include mechanical shutters, electro-optic shutters, acousto-optic shutters, and passive shutters. Mechanical shutters are briefly outlined in the following paragraphs.

With mechanical shutters, one of the cavity mirrors, usually the total reflector, is rotated about an axis perpendicular to the resonator axis (Figure 1.31). Just before the two mirrors become parallel, the flashlamp is triggered to initiate pumping action. However, since the mirrors are then not aligned (closed shutter case), the cavity losses are too high to sustain oscillation. The population inversion is thus increased beyond the threshold value. The flashlamp trigger is timed such that the population inversion will be a maximum just when the two mirrors are parallel. The energy stored in the form of high population inversion is then released as a high-power pulse when the mirrors become aligned (open shutter case).

The mechanical shutter technique is robust and wavelength-independent. However, the normal mirror rotational speeds of up to 30,000 revolutions per minute (RPM) result in slow switching action, which in turn results in a relatively lower peak power. Furthermore, the high rotational speeds may result in mechanical vibration that can influence the beam stability. Also, obtaining ideal beam alignment is a significant problem. The vibrational problem may be minimized by using a quadrilateral- or hexagonal-shaped mirror assembly. As a result, it is difficult to use mechanical shutters for Q-switching, especially for lasers of very short upper-level lifetime, such as argon ion lasers.

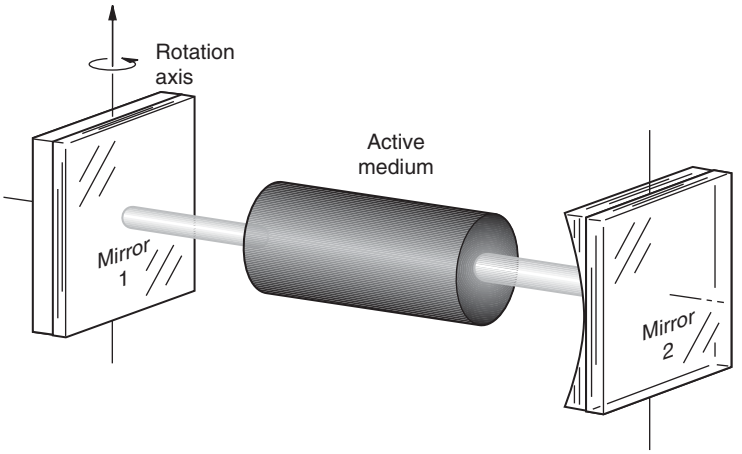


Figure 1.31 Mirror rotation for mechanical shutter action.

The typical pulse duration obtained when using mechanical shutters is of the order of 400 ns. Increasing the rotational speed would increase the peak power and reduce the pulse duration. They are not extensively used because of the mechanical problems associated with them.

Electro-optic shutters are usually not as robust as mechanical shutters but are much faster. They are normally useful for single-pulse operation. Acousto-optic modulators are normally used when pumping is continuous to produce repeated pulsing, such as is done with neodymium and argon ion lasers.

Passive Q-switching is based on saturable dyes or absorbers. They are normally used for single-pulsed operation.

1.6.3 Mode Locking

Mode locking is used to generate pulses of higher peak power (of the order of gigawatts or more) than can be achieved by Q-switching, and of very short duration (of the order of picoseconds or femtoseconds).

A laser cavity generally sustains a large number of oscillating modes. The oscillation of each of these modes is normally independent of the other modes. Mode locking is achieved by combining a number of distinct longitudinal modes of a laser in phase, with each mode having a slightly different frequency.

In the ideal case, the various longitudinal modes in a planar resonator are equally spaced (Figure 1.12c), with a frequency separation given by:

$$\delta\omega = \pi\frac{c}{L} \text{ or } \delta\nu = \frac{c}{2L} \tag{1.59}$$

where L is the cavity length.

However, nonlinear dispersion of the active medium results in the various modes having random phases, ϕ_{m}, with the frequency differences between adjacent modes also being random. The result is a number of irregular spikes in the laser output. The total electric field, $E_{\mathrm{l}}(t)$, of the corresponding electromagnetic wave is a superposition of the fields of all modes and can be expressed as:

$$E_{\mathrm{l}}(t) = \sum_{m=1}^{M_{\mathrm{c}}} E_{0\mathrm{m}} e^{i(\omega_{\mathrm{m}}t+\varphi_{\mathrm{m}})} \tag{1.60}$$

where E_{0m} is the amplitude of oscillation of mode m, ω_m is the frequency of mode m, ϕ_m is the phase, and M_c is the number of oscillating modes.

For simplicity, let us consider the case where the various modes are of equal amplitude, E_0. The overall peak beam intensity is then a summation of the intensities of the individual modes:

$$I_p = M_c {E_0}^2 \tag{1.61}$$

Let us now consider the case where the individual modes are locked in phase (forced to maintain fixed-phase relationships) such that in addition to having the same amplitude, they also oscillate with equal relative phase:

$$\varphi_m = \varphi \tag{1.62}$$

If the separation between modal frequencies, $\delta\omega$, are also equal, then we have

$$\omega_m = \omega + m\delta\omega \tag{1.63}$$

where ω is the frequency of the lowest mode.

The electric field can then be expressed as:

$$E_l(t) = E_0 \sum_{m=1}^{M_c} e^{i[(\omega+m\delta\omega)t+\varphi]} = E_0 e^{i(\omega t+\varphi)} \sum_{m=1}^{M_c} e^{im\delta\omega t} \tag{1.64}$$

which when expanded becomes

$$E_l(t) = E_0 e^{i(\omega t+\varphi)}[e^{i\delta\omega t} + e^{i2\delta\omega t} + e^{i3\delta\omega t} + \cdots + e^{iM_c\delta\omega t}] \tag{1.65}$$

The overall beam intensity, which is proportional to $|E_l(t)|^2$, is then given as:

$$I(t) = {E_0}^2[e^{i\delta\omega t} + e^{i2\delta\omega t} + e^{i3\delta\omega t} + \cdots + e^{iM_c\delta\omega t}]^2 \tag{1.66}$$

which forms a geometric series that reduces to the form:

$$I(t) = {E_0}^2 \frac{\sin^2\left(\dfrac{M_c\delta\omega t}{2}\right)}{\sin^2\left(\dfrac{\delta\omega t}{2}\right)} \tag{1.67}$$

The peak intensity, I_p, is obtained by considering small values of time when $t \approx 0$ and thus $\sin\theta \approx \theta$, giving

$$I_p = {M_c}^2 {E_0}^2 \tag{1.68}$$

This is the peak intensity of the mode-locked laser which is thus M_c times that of the nonmode-locked laser. For a typical solid-state laser where the number of oscillating modes may be very large, of the order of 10^4, this can be very significant. The average power, however, is unaffected by mode locking. The peak intensity occurs at times when $\sin \delta\omega t/2 = 0$:

$$\frac{\delta\omega t}{2} = m\pi, \qquad m = 1, 2, 3, \ldots \tag{1.69}$$

or

$$t = \frac{2m\pi}{\delta\omega} = \frac{2mL}{c} \tag{1.70}$$

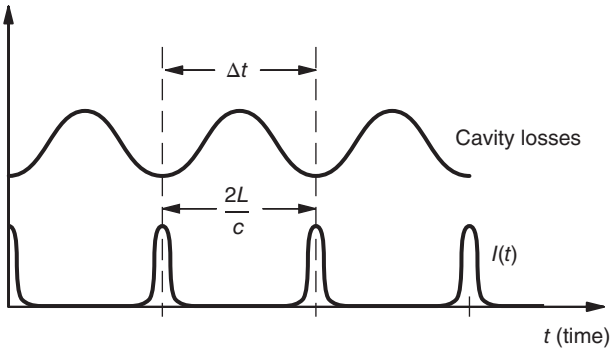


Figure 1.32 Amplitude modulation of cavity losses. Source: From Svelto (2010)/Springer Nature.

The separation between contiguous pulses is then given by:

$$\Delta t = \frac{2L}{c} \tag{1.71}$$

which is the time taken for a round trip of the cavity. The mode-locked laser may thus be considered as constituting a pulse that propagates back and forth in the cavity. It can be shown that the pulse duration, $\Delta\tau_p$, (FWHM) of the beam is given by:

$$\Delta\tau_p \approx \frac{2\pi}{M_c\delta\omega} = \frac{2\pi}{\Delta\omega_0} \tag{1.72}$$

Thus, the larger the oscillating bandwidth, $\Delta\omega_0$, the shorter the output pulse duration. For gas lasers, the pulse duration is typically of the order of nanoseconds, whereas for solid-state lasers, the pulse duration may be of the order of picoseconds or less.

The common techniques used for mode locking may be categorized as either active or passive. We briefly outline the active mode locking method. With this method, an external signal is used to modulate or vary cavity losses or quality factor, $Q_f(t)$, at a frequency equal to the intermodal separation (Figure 1.32). This can be done using either electro-optic or acousto-optic shutters. We consider the simple case where the shutter is open very briefly, $\delta\tau$, after every period of $\Delta t = 2L/c$, and closed the rest of the time ($\Delta t - \delta\tau$) during that period. A pulse of light of exactly the same length as the time $\delta\tau$ that the shutter stays open, and which arrives exactly when the shutter is open, will then be unaffected by the shutter. Any portion of the signal that arrives when the shutter is closed will be eliminated. A single pulse that propagates back and forth thus results.

Mode locking induced by cavity losses as described is said to be amplitude-modulated. It can also be induced by variation in the refractive index of the cavity, which results in variations in optical path length. It is then said to be frequency-modulated.

1.7 BEAM CHARACTERISTICS

The properties of a laser beam that have significant influence on materials processing include divergence, monochromaticity, coherence, brightness, stability, size, and mode.

1.7.1 Beam Divergence

Beam divergence may be defined as the characteristic property of a light beam that results in an increase in its cross-sectional area with distance from the source. Light from an ordinary source is uniformly radiated in all directions, and as a result, its divergence is high (Figure 1.33a). Its intensity is inversely proportional to the square of the distance from the source. A laser, on the other hand, is such that it is only light which is almost parallel with the resonator axis, and which is close to the axis that comes out, resulting in relatively low divergence (Figure 1.33b).

Divergence of a diffraction-limited monochromatic beam of wavelength λ that consists of plane waves which are incident on a circular aperture of diameter D can be expressed, from diffraction

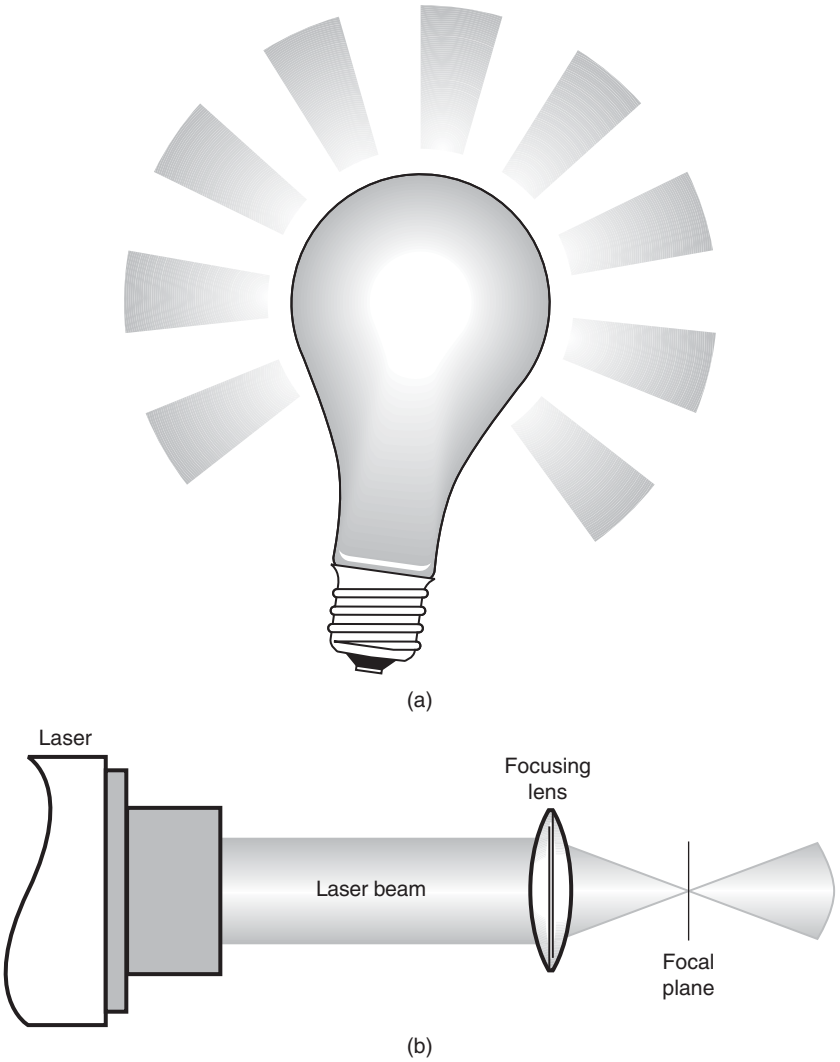


Figure 1.33 (a) Divergent light from an ordinary light source. (b) Light from a laser beam showing minimal divergence.

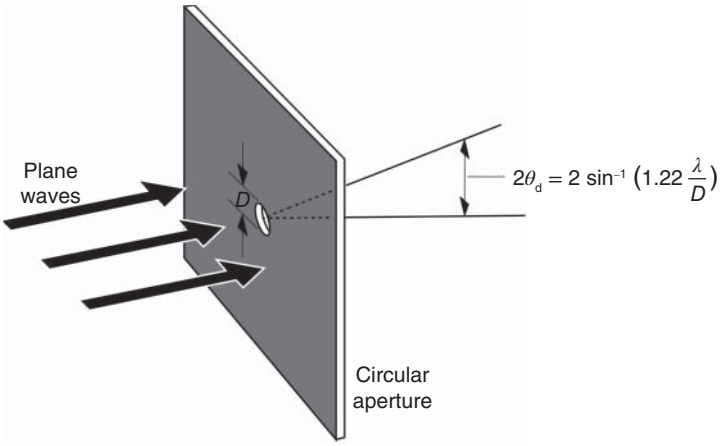


Figure 1.34 Beam divergence. Source: From O'Shea et al. (1977)/with permission of Pearson.

theory, by the semiangle θ_d as (Figure 1.34):

$$\theta_d = \sin^{-1}\left(\frac{K_1\lambda}{D}\right) \approx \frac{K_1\lambda}{D} \tag{1.73}$$

where K_1 is a constant ≈1.22.

The low divergence of a laser enables it to be focused to a very small radius. Single-mode lasers generally experience less divergence than multimode lasers. Also, a Gaussian beam exhibits the smallest possible beam divergence (true diffraction-limited beam). A typical value for the divergence, θ_d, may be 0.05°. However, for semiconductor lasers, the divergence can be much higher, by about two orders of magnitude (Appendix 2.B).

For some manufacturing processes such as laser cutting and microfabrication, it is essential to have a highly focused beam. Thus, low beam divergence is important. It may not be as important for high-power laser welding, and even less important for surface heat treatment or laser forming, where a wider beam area is more desirable.

1.7.2 Monochromaticity

A light beam is said to be monochromatic when it is composed of only a single wavelength. Lasers are generally considered to be monochromatic (Figure 1.35). However, in reality, due to the broadening mechanisms discussed earlier, the wavelength of a laser is not exactly monochromatic but covers a frequency bandwidth. For lasers, the oscillating bandwidth can be very narrow and may range from $\Delta\nu = 500$ Hz ($\Delta\lambda = 10^{-11}$ μm at $\lambda = 0.6$ μm; Section 3.1) for a high-quality stable gas laser to greater than 1 GHz ($\Delta\lambda = 10^{-6}$ μm) for a solid-state laser. In comparison, the bandwidth for a conventional monochromatic source such as the sodium lamp may be of the order of several gigahertz.

Theoretically, the minimum oscillating linewidth (or fluctuation of the laser frequency) $\delta\nu$ of a laser oscillating in a single mode under steady state is determined by spontaneous emission and can be shown to be given by:

$$\delta\nu = \frac{4\pi h_p \nu_0 (\Delta\nu)^2}{q} \tag{1.74}$$

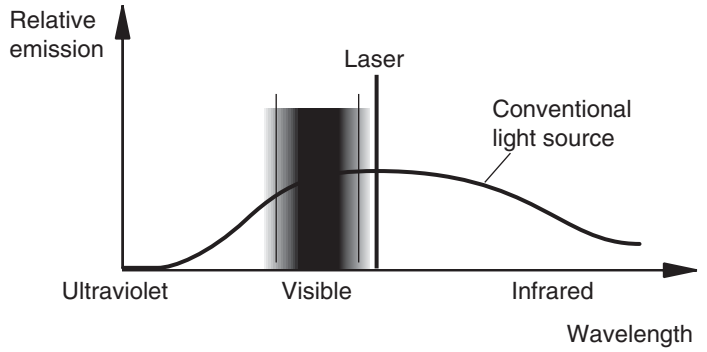


Figure 1.35 Comparison of the spectral characteristics of a conventional light source and a laser source.

where q is the power in the oscillating field, h_p is Planck's constant, $\Delta\nu$ is the half-width of the resonance at half maximum intensity, ν_0 is the oscillation frequency, and $\delta\nu$ is the minimum oscillating linewidth.

Equation (1.74) indicates that an increase in the laser power results in a decrease in the theoretical linewidth. This is because an increase in power is due to an increase in the ratio of stimulated to spontaneous emissions.

For a pulsed laser beam, the minimum linewidth is given by the inverse of the pulse duration, $\Delta\tau_p$:

$$\delta\nu = \frac{1}{\Delta\tau_p} \tag{1.75}$$

Monochromaticity is important in laser processing, because it affects beam focusing. A monochromatic beam can be more sharply focused than a beam with a broad bandwidth. This may not be significant in laser welding or heat treatment, but it is important in laser cutting or microfabrication.

1.7.3 Beam Coherence

Coherence refers to a fixed-phase relationship between two waves or between two points of the same wave (Figure 1.36a). It is both space (spatial)- and time (temporal)-dependent and is important for interference-based measurements.

Spatial coherence is where the phase difference between two points on a wavefront of an electromagnetic field remains constant with time (Figure 1.36a,b). Temporal coherence, on the other hand, is where the phase difference between the wavefront of an electromagnetic field at a given point P at a time t and that at the same point P and time $t+\tau_0$ remains constant with time. The electromagnetic wave is then temporally coherent over the period τ_0. This is illustrated in Figure 1.36b, which shows an electromagnetic wave whose phase changes at intervals of τ_0. Even though a beam may have both spatial and temporal coherence, the two do not necessarily coexist. For clarity, an incoherent beam is illustrated in Figure 1.36c.

1.7.3.1 Spatial Coherence The concept of spatial coherence is best understood using Young's experiment. A light source S is placed in front of a narrow slit A_0 contained in a screen A. We set up a coordinate system as shown in Figure 1.37, with origin O at the center of the slit.

A second screen B containing two parallel and narrow slits B_1 and B_2 is placed at a distance x_1 from A such that the y-axis is exactly halfway between the two slits B_1 and B_2 which are parallel

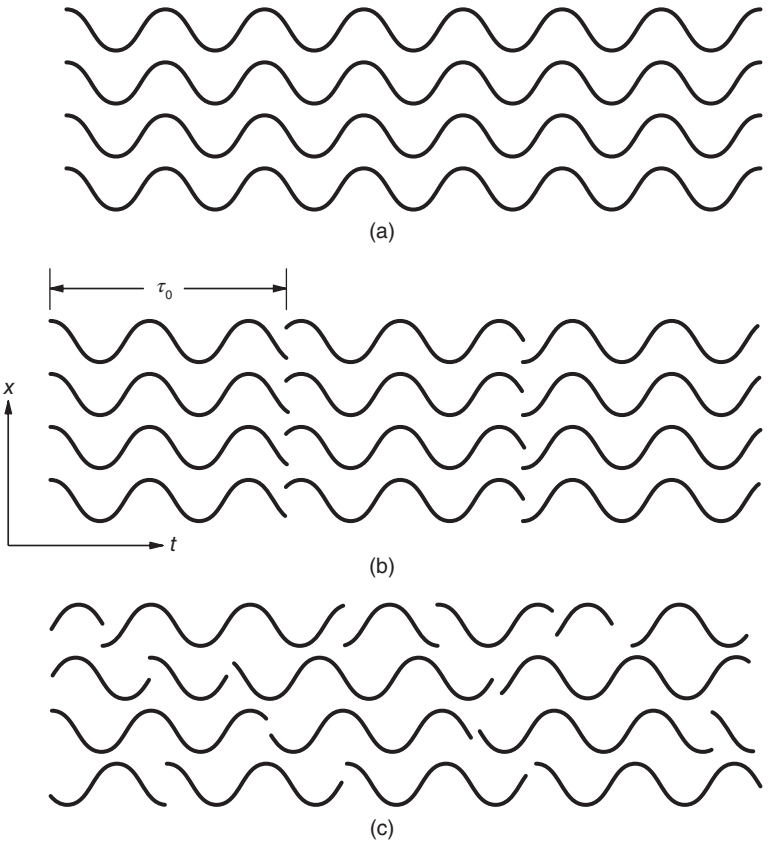


Figure 1.36 Coherence: (a) perfect coherence. (b) Spatially coherent beam with only partial temporal coherence with coherence time τ_0. (c) Almost completely incoherent beam. Source: From Wilson and Hawkes (1987)/with permission of Pearson.

to the y-axis and are located at a distance $\pm z_b$ from the x-axis. A third screen C is now placed at a distance x_2 from B such that the x-axis passes through it at D. The three screens are all parallel to the y–z plane.

It can be shown that light from the origin O and passing through the slits B_1 and B_2 will reinforce each other if the path difference, l, at $D'(x_1 + x_2, z)$ is an integer number of wavelengths:

$$l = \frac{2zz_b}{x_2} = m\lambda \tag{1.76}$$

and they will nullify each other if the path difference is an integer number of wavelengths and a half:

$$l = \frac{2zz_b}{x_2} = (m + 1/2)\lambda \tag{1.77}$$

Thus, as D′ varies with distance from D, the light intensity will show alternate bright and dark fringes. From Eq. (1.76), the position of the mth bright fringe can be obtained as:

$$z(m) = \frac{mx_2\lambda}{2z_b} \qquad m = 0, \pm 1, \pm 2, \ldots \tag{1.78}$$

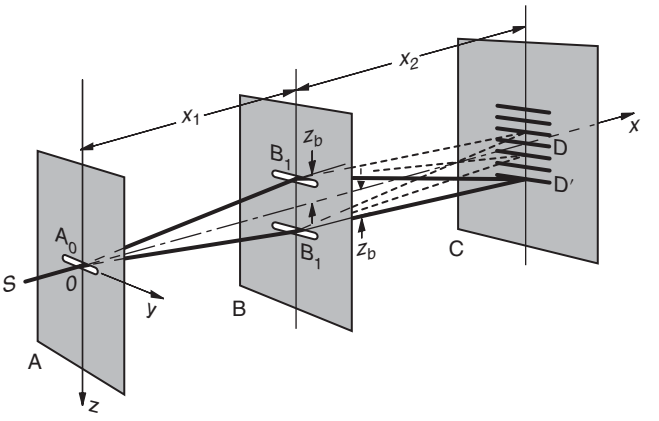


Figure 1.37 Young's experiment.

The spacing between contiguous fringes will be

$$z(m+1) - z(m) = \frac{x_2 \lambda}{2 z_b} \tag{1.79}$$

If now another slit is placed in screen A which is parallel to the one at O and located at O′, a distance d_s from O and is very close to O, then the fringes formed by O′ will be such that there will be a bright fringe in the neighborhood of D. If the distance d_s is now gradually increased by moving O′ in the positive z direction, its fringes will shift relative to those due to O. At some critical separation d_c, the dark fringes of O′ will be superimposed on the bright fringes of O, eliminating the entire interference fringe pattern on the screen. d_c can be shown to be given by:

$$d_c = \frac{x_1 \lambda}{4 z_b} \tag{1.80}$$

If O and O′ are considered to be part of a relatively wide single slit of width d_s, then it can be said that good fringes will be obtained if

$$d_s \ll \frac{x_1 \lambda}{z_b} \Rightarrow z_b \ll \frac{x_1 \lambda}{d_s} \tag{1.81}$$

Thus, for a wide enough slit, no fringes will be observed for an ordinary light source, and this is because the waves emitted from different parts of the source are incoherent. A laser beam, on the other hand, will produce interference fringes even for a wide slit since it has high spatial coherence, and thus light from any two points on the beam can interfere.

1.7.3.2 Temporal Coherence The concept of temporal coherence is also best explained using the Michelson interferometer, (Figure 1.38). A parallel beam of near-monochromatic light from a source S is split into two using a beam splitter A, such that part of the beam is reflected and the other part transmitted. Each of the two beams is reflected back from a plane mirror. One of the mirrors, M_1, is fixed, while the other, M_2, is movable along the direction of the beam. Each of the reflected beams is again partly reflected and partly transmitted at A such that they are both directed onto a detector at B where they are superimposed to form interference fringes. If the distance from A to M_1 is the

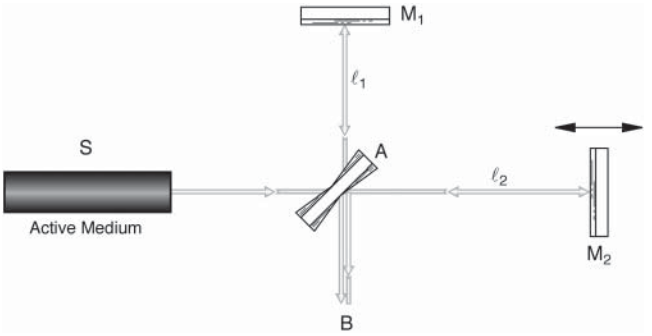


Figure 1.38 Michelson interferometer.

same as that from A to M_2, then the optical path difference between the reflected beams is zero. Good interference fringes are then observed at B. On gradually moving the mirror M_2 away from A, the interference fringes become less and less distinct, until they completely disappear when the optical path difference is a few centimeters (Figure 1.39).

The reduction in fringe visibility with increasing optical path difference is due to the fact that the light waves generated by the source are not of an infinitely long duration, but rather a series of wave trains of average duration say, τ_{ct}. The phases of the individual wave trains are random. Despite the randomness of the wave trains, a light source appears continuously bright because the wave trains are being constantly produced and superimposed on each other.

The duration of a wave train, τ_{ct}, is referred to as the coherence time, and its length $c\tau_{ct}$ is the coherence length. The finite duration or length of the wave trains of an ordinary light source is the reason why the fringes produced by the Michelson interferometer reduce in quality as the optical path difference between the two beams increases. When the optical path difference is less than the coherence length, then light waves are being superimposed which were essentially split from the same wave train and are thus in phase. Good interference fringes are obtained between light waves which are coherent. When the optical path difference is greater than the coherence length, then different light waves are being superimposed which were generated at times separated by an interval greater than the coherence time and are thus uncorrelated.

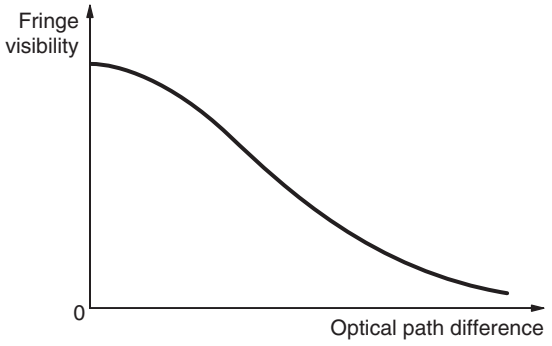


Figure 1.39 Variation of fringe visibility with optical path difference. Source: From Shimoda (1986)/Springer Nature.

The linewidth $\Delta\nu_0$ associated with a light wave having a coherence time of τ_{ct} can be shown to be given by:

$$\Delta\nu_0 = \frac{1}{\tau_{ct}} \tag{1.82}$$

The linewidth associated with the emission from an ordinary source with a coherence time of about $\tau_{ct} = 10^{-10}$ s will thus be 10^{10} Hz. On the other hand, the oscillating linewidth for a high-quality stable gas laser may be as small as 500 Hz. The corresponding coherence time and length are 2×10^{-3} s and 6×10^5 m (600 km), respectively. Thus, the interference properties of a laser are very good and can be used for producing interference fringes even when the optical path difference is large.

1.7.4 Intensity and Brightness

The brightness of a power source at a point is defined as the power emitted per unit surface area per unit solid angle (angular cone). Brightness is a basic characteristic of the source radiation and cannot be increased. The brightness of different light sources may be more easily compared by specifying the brightness temperature, T_B, which is defined as that temperature at which a blackbody would emit the same radiant flux as the given body and may be expressed as:

$$T_B = \frac{q}{k_B \Delta\nu_0} \tag{1.83}$$

where q is the power of the source, $\Delta\nu_0$ is the linewidth, and k_B is Boltzmann's constant.

For a 1 W laser of linewidth of 1 MHz, the corresponding brightness temperature is of the order of 10^{19} K. For comparison, we note that the brightness temperature of an incandescent lamp is about 10^4 K.

1.7.5 Focusing

For an almost parallel Gaussian beam of wavelength λ incident on a lens, the radius w_f at the focus is given (Figure 1.40) by:

$$w_f = \frac{\lambda}{\pi\theta} \approx \frac{1}{\pi}\frac{f_l \lambda}{w_0} \tag{1.84}$$

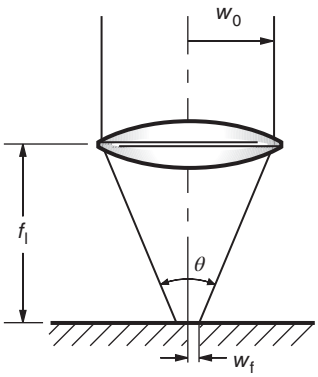


Figure 1.40 Beam focusing.

where f_l is the lens focal length, w_0 is the beam radius at the lens, and θ is the angular aperture of the lens, or the angle subtended by the lens at the focus.

The minimum spot diameter, d_{min}, that can be obtained at the focal point is given by:

$$d_{min} = 2w_{min} = \frac{1.22\lambda}{NA} \approx \lambda \tag{1.85}$$

where NA $= n \sin(\theta/2)$ is the numerical aperture, and n is the refractive index of the medium in which the lens is working.

For traditional welding and cutting applications, the beam diameter at the focal point is about 250 μm. However, with appropriate optics, this can be reduced to the order of the beam's wavelength.

For a general mode distribution, the radius at the focus may be expressed as:

$$w_f = K_m \frac{1}{\pi} \frac{f_l \lambda}{w_0} \tag{1.86}$$

where K_m represents the energy mode distribution. A typical value for K_m would be 1.

It is evident from Eq. (1.84) that the smaller the wavelength of the laser beam, the smaller the focused beam size, and thus the greater the power density produced by a beam of given power. As a result, for a given beam power and mode, the power density produced with an Nd:YAG laser will be two orders of magnitude greater than that produced using a CO_2 laser.

1.8 SUMMARY

A laser is a form of electromagnetic radiation with wavelengths ranging from X-ray to infrared radiation. It is the result of energy emission associated with the transition of an electron from a higher to a lower energy level within an atom. This starts with excitation of an atom or molecule to a higher energy state as it absorbs a photon. When the excited atom is stimulated by a photon, it also releases another photon as it undergoes a transition to the lower state. This results in stimulated emission, where the incident and emitted photons have the same characteristics and are in phase, resulting in a high degree of coherence. Otherwise, the photon is released spontaneously. Under thermal equilibrium conditions, the distribution of atoms at the various energy levels is given by Boltzmann's law.

Laser action is achieved by placing the active medium between two mirrors to form the resonator. Simple resonator forms include the planar, confocal, and concentric resonators. The major mode forms that are sustained in a resonator are the longitudinal (axial) and transverse modes. Contiguous longitudinal modes only differ in their field distribution along the longitudinal axis, while transverse mode field distributions differ in the transverse direction. Undesirable modes can be suppressed using mode selection methods. Higher-order transverse modes may be suppressed by placing a diaphragm in the cavity. Longitudinal mode selection methods include cavity length variation.

Pumping is the process by which atoms are raised from a lower to a higher energy level through stimulated absorption. This is done to achieve population inversion. Two common forms of pumping are optical and electrical pumping.

Optical pumping can be done using either lamp-based or diode laser-based systems. Arc and flashlamps are normally used with solid-state or liquid lasers, since the lamp output usually has a broad bandwidth. Configurations that are used when pumping with diode lasers are longitudinal (end) and transverse (side) pumping. Electrical pumping is normally done using radio frequency or direct current systems and is more commonly used with gas and semiconductor lasers.

Population inversion is difficult to achieve in a two-level system by pumping. It is easier for a three-level system, where the lifetime for transition from level 3 to level 2 must be smaller than

the lifetime for transition from level 2 to level 1. For a four-level system, population inversion can be achieved when the lifetime for transition from level 1 to level 0 is smaller than the lifetime for transition from level 2 to level 1. Thus, level 2 is essentially in a metastable state, while levels 3 and 1 have relatively short lifetimes.

Even though lasers are generally considered to be monochromatic, in reality, the transition frequency is not exactly monochromatic, but it is spread over a finite frequency range. The two major categories of line-broadening are homogeneous and inhomogeneous broadening. Under homogeneous broadening, we have natural and collision broadening. Natural broadening is associated with the finite lifetime of the excited state as a result of spontaneous emission. Collision broadening results primarily from collision between atoms and is therefore more predominant in gaseous media.

In homogeneous broadening, the mechanism has the same effect on each atom, and thus the response of each atom is broadened in an identical manner. The entire system is thus broadened in the same way. Inhomogeneous mechanisms include Doppler broadening, which is a result of the Doppler effect.

The Q-factor, Q_f, of a laser cavity is a measure of the losses or energy dissipation in any mode of the laser cavity. A cavity with a high Q-factor has low losses and thus has a smaller linewidth.

Q-switching is a technique that is used to produce pulsed laser outputs of higher power but shorter duration. It involves using a shutter to control the Q-factor by preventing oscillation in the cavity until a high population inversion is achieved, at which time the shutter is opened, resulting in an intense beam of short duration being released. Q-switching can be achieved using mechanical, electro-optic, acousto-optic, or dye (passive) shutters.

Pulses with higher peak powers and shorter duration than can be obtained by Q-switching are achieved through mode locking. This is achieved by combining a number of distinct longitudinal modes of a laser in phase. Mode locking is achieved using electro-optic or acousto-optic shutters.

The divergence of a laser beam is proportional to the beam wavelength and inversely proportional to the diameter of the aperture from which the beam is emerging. Whereas ordinary light is uniformly radiated in all directions, the divergence of a laser beam may be as low as 0.05°. The size of a focused beam is proportional to the wavelength of the incident beam and inversely proportional to the size of the beam that is incident on the lens.

APPENDIX 1.A

List of symbols used in the chapter

Symbol	Parameter	Units
$2a \times 2b \times L$	Dimensions of rectangular cavity	m
A_e	Einstein coefficient for spontaneous emission	/s
B_{12}	Einstein coefficient for stimulated absorption	Sr m^2/J s
B_{21}	Einstein coefficient for stimulated emission	Sr m^2/J s
$B = B_{12} = B_{21}$	Einstein coefficient	Sr m^2/J s
c	Speed of light	m/s
c_m	Velocity of light in the active medium	m/s
d_c	Critical separation between slits	m
d_s	Separation betwen slits	m
d_{min}	Minimum focused beam diameter	μm
D	Aperture diameter	m
$e(\nu)$	Energy density (energy per unit volume) at the frequency ν	J/m^3 Hz

Symbol	Parameter	Units
E_{0m}	Amplitude of oscillation of mode m	V/m
FWHM	Full width at half maximum	rad/s
F_i	Focal points	—
f_l	Lens focal length	mm
$g(\omega, \omega_0)$	Line shape function	—
h_i	Distance of mirror M_i from beam splitter	m
h_p	Planck's constant $= 6.625 \times 10^{-34}$	J s
I	Beam intensity	W/cm^2
I_p	Peak beam intensity	W/cm^2
$I(\omega)$	Power spectrum of intensity	W/m^2
K_1	Constant	—
k_B	Boltzmann's constant $= 1.38 \times 10^{-23}$	J/K
K_m	Energy mode distribution	—
k_s	Spring constant	N/m
k_x	$= 2\pi/\lambda =$ wavenumber	/m
l	Optical path length difference between two rays	m
m, n, p	Integers	—
M_c	Number of modes in closed cavity	—
M_i	Mirror i	—
m_m	Mass of atom or molecule	mg
n_{abs}	Absorption rate (number of absorptions/unit volume/unit time)	/m^3 s
n_{sp}	Spontaneous emission rate	/m^3 s
n_{st}	Stimulated emission rate	/m^3 s
NA	Numerical aperture	—
P_g	Gas pressure	N/m^2
P_l	Electric polarization	C m^2
q	Power	W
q_l	Power loss in a mode	W
Q	Energy stored in a mode	J
Q_0	Initial energy at time $t = 0$	J
Q_f	Quality factor	—
r	Radial position	cm
r_a	Radius of atom or molecule	μm
S_{ij}	Probability per unit time of non-radiative decay between levels i and j	/s
R_i	Reflectivity of mirror i	—
Δt	Time separation between contiguous pulses	s
t_{cb}	Cavity buildup time	s
t_c	Time between collisions	s
t_{ca}	Average time between collisions	s
t_p	Pumping duration	s
T	Absolute temperature	K
T_B	Brightness temperature	K
x_1	x-coordinate of plane B	m
x_2	Distance between planes B and C	m
z_b	z-coordinate of slits in plane B	m
u_x	Velocity component in the x-direction	m/s

Symbol	Parameter	Units
V_c	Volume of closed cavity	m^3
w	Beam radius	μm
w_0	Beam radius at the lens	μm
w_m	Beam spot size on the mirror	μm
w_{min}	Minimum focused beam radius	μm
w_w	Beam radius at the waist	μm
W_{ij}	Probability per unit time of the atom being thermally excited between levels i and j	/s
x	Displacement of electron from equilibrium position in x-direction	m
X_0	Amplitude of oscillation	m
x_R	Rayleigh range	m
α	Absorption coefficient	/m
β	Small-signal gain coefficient	/m
β_{th}	Small-signal threshold gain coefficient	/m
ϕ_m	Phase of mode m	o
γ_0	Damping coefficient	/s
λ	Wavelength	μm
ν	Frequency	Hz
ω	Angular velocity	rad/s
ω_0	Resonant or transition frequency	rad/s
ω_m	Frequency of mode m	rad/s
$\delta\omega$	Frequency separation between longitudinal modes	rad/s
θ_d	Divergence semiangle	°
τ_0	Coherence period	s
τ_c	Cavity lifetime	s
τ_{ct}	Coherence time	s
τ_{ij}	Lifetime for transition from level i to level j	s
τ_{sp}	Lifetime	s
τ_u	Upper-level lifetime	s
$\Delta\tau_p$	Pulse duration	s
$\delta\tau$	Shutter opening time	s
$\Delta\tau_p$	Pulse duration	s
$\delta\nu$	Minimum oscillating linewidth	Hz
$\Delta\nu_0$	Linewidth	Hz
χ	Electric susceptibility	—

PROBLEMS

1.1. Consider the energy levels E_1 and E_2 of a two-level system. Determine the population ratio of the two levels if they are in thermal equilibrium at room temperature, 27 °C, and the transition frequency associated with this system is at 10^{15} Hz.

1.2. For the system in Problem 1.1, determine the fraction of atoms that will be in the lower state at room temperature, 27 °C, if the transition wavelength associated with the two energy levels is (a) $\lambda = 1060$ nm and (b) $\lambda = 488$ nm.

1.3. If the population ratio, N_2/N_1, for the energy levels in Problem 1.1 under thermal equilibrium conditions at room temperature is $1/e^2$, determine the transition frequency and the type of electromagnetic radiation associated with this transition.

1.4. The oscillating wavelengths of the He–Ne, Nd:YAG, and CO_2 lasers are 0.6328, 1.06, and 10.6 μm, respectively. Determine the corresponding oscillating frequencies. What energy is associated with each transition?

1.5. Determine the stimulated emission probability, B, for a transition of wavelength $\lambda = 250$ nm, if its spontaneous emission rate, A, is 5×10^5/s. For the stimulated emission probability to be 400% that of spontaneous emission, what must be the irradiance at the given wavelength within the cavity at room temperature, 27 °C?

1.6. A medium absorbs 0.5% of the light that passes through it for each millimeter of medium length. Determine its absorption coefficient, and the percentage of light that will be transmitted if the overall length of the medium is 120 mm.

1.7. (a) Neglecting any losses, determine β for an active medium of length 2 m if the irradiance of light passing through it increases by 100%.
(b) Determine β if a 20% increase in irradiance occurred for the same length of active medium.

1.8. Determine the radiation density emitted by a blackbody at a frequency of 10^{15} Hz, if the temperature is 1200 K, using Planck's theory.

1.9. Consider a transition of wavelength $\lambda = 300$ nm. Obtain the stimulated emission probability, B, if the spontaneous emission rate is 4×10^5/s. Further, define the cavity in terms of the energy density at room temperature (298 K) if the stimulated emission probability is 500% of the spontaneous emission probability.

1.10. A laser cavity has a wavelength centered at $\lambda = 500$ nm. If the cavity volume is 3500 mm^3, determine the number of modes that fall in a bandwidth of $\Delta\lambda = 0.3$ nm.

1.11. Using a power series expansion, derive the following equation for a rectangular cavity where $a = b$:

$$\nu = \frac{\omega}{2\pi} = \frac{c}{2}\left[\frac{p}{L} + \left(\frac{m^2 + n^2}{a^2}\right)\frac{L}{8p}\right]$$

given that

$$\omega^2 = c^2\pi^2\left(\frac{m^2}{4a^2} + \frac{n^2}{4b^2} + \frac{p^2}{L^2}\right)$$

1.12. Given that

$$\omega^2 = c^2\pi^2\left(\frac{m^2}{4a^2} + \frac{n^2}{4b^2} + \frac{p^2}{L^2}\right)$$

and that p is constant, $m = n$, and m changes by 1, that is, $\Delta m = 1$, show that the corresponding change in frequency is given by:

$$\Delta\nu_{\mathrm{m}} = \frac{cL}{8a^2p}\left(m + \frac{1}{2}\right)$$

1.13. Given

$$\Delta\nu_{\mathrm{m}} = \frac{cL}{8a^2p}\left(m + \frac{1}{2}\right)$$

show that

$$\Delta\nu_{\mathrm{m}} \approx \Delta\nu_{\mathrm{p}}\frac{\lambda L}{8a^2}\left(m + \frac{1}{2}\right)$$

1.14. If the mirrors of a planar resonator are placed at a distance of 30 cm apart, and the mirror dimensions are $a = b = 10$ mm, determine

(a) The transmitted wavelengths that will be immediately next to the 0.633 μm wavelength.
(b) The m value for that wavelength.
(c) The wavelength of the next integer mode transmission.
(d) The frequency spacing between two contiguous transverse mode transmissions.

1.15. For a confocal resonator, show that the beam spot size at any general location along the cavity axis is also given by:

$$w(x) = w_{\mathrm{w}}\left[1 + \left(\frac{\lambda x}{\pi {w_{\mathrm{w}}}^2}\right)^2\right]^{1/2}$$

1.16. The Gaussian field distribution for a confocal resonator is given by:

$$g_{00}(y, z) = e^{-\frac{\pi}{L\lambda}(y^2+z^2)}$$

show that the beam spot size of the electric field amplitude on the mirrors is

$$w_{\mathrm{m}} = \left(\frac{L\lambda}{\pi}\right)^{1/2}$$

1.17. Given that

$$w(x) = w_{\mathrm{w}}\left[1 + \left(\frac{2x}{L}\right)^2\right]^{1/2}$$

show that the following relationship holds

$$w(x) \approx \frac{\lambda x}{\pi w_{\mathrm{w}}} \qquad \text{as} \quad x \to \infty$$

1.18. A CO_2 laser with output wavelength 10.6 μm has a cavity of length 1500 mm and a confocal mirror system.

(a) What is the beam spot size at the resonator center?
(b) What will be the spot size on the mirrors?
(c) What will be the frequency difference between two contiguous longitudinal modes?

1.19. Given that the beam size from a confocal resonator is given by:

$$w_{\mathrm{w}} = \left(\frac{L\lambda}{2\pi}\right)^{1/2}$$

and

$$w(x) = w_w\left[1 + \left(\frac{2x}{L}\right)^2\right]^{1/2}$$

show that the beam divergence may be approximated by:

$$\theta \approx \frac{\lambda}{\pi w_w}$$

1.20. Consider a cylindrical tube of radius r and length L.

(a) If a Gaussian beam propagates through the cylinder, with its beam waist at the tube center, obtain an expression for the beam waist when the transmittance is maximum, assuming the power that is incident at the entrance of the cylinder can be transmitted.

(b) If the cylinder diameter is 5 mm, and the length is 100 mm, determine the fraction of incident He–Ne laser Gaussian beam that will be transmitted.

1.21. If the life of a cavity is τ_c, show that the time constant of the resulting laser light amplitude is $2\tau_c$.

1.22. Why does the Gaussian beam exhibit the smallest possible beam divergence?

1.23. A laser beam would normally contain a number of axial modes. What mechanism could be used in the generation of the beam to produce only a single axial mode?

1.24. Show that the Rayleigh range, x_R, is given by:

$$x_R \approx \frac{\pi {w_w}^2}{\lambda}$$

1.25. Two types of laser beam (A and B) which are based on different active media are generated using identical confocal resonators. The spot sizes of the beams at the waist are 1.3 and 0.41 mm, respectively, and the spot sizes at a specific distance from the waist are 2.03 and 0.64 mm, respectively. Calculate

(a) The distance separating the confocal mirrors.
(b) The wavelength of the lasers.
(c) The Rayleigh range of both lasers.
(d) The spot size at the Rayleigh range for each laser.

1.26. For confocal resonators, the radius of the beam, $w(x)$, is given by Eq. (1.30). Show that this reduces to the form:

$$w(x) \approx \frac{\lambda x}{\pi w_w} \text{ as } x \to \infty$$

(a) Characterize the error in using the approximation by using a solver to plot the two functions from $x = 0$ mm to $x = 1000$ mm. Use the following conditions: $L = 1500$ mm, $\lambda = 10.6$ μm.

(b) Plot the error as a percentage of Eq. (1.30).

(c) Using plots, what is the range of x values that result in the error being less than 20%, 10%, and 5%?

1.27. Show that the full width at half maximum of natural broadening is given by:

$$\Delta\omega_0 = 2\Delta\omega = \frac{1}{\tau_{sp}}$$

1.28. Given that the power spectrum of the radiation due to random collisions is given by:

$$f(\omega) = \int_0^\infty \frac{{X_0}^2}{\pi^2} \frac{\sin^2\left[(\omega-\omega_0)\frac{t_c}{2}\right]}{(\omega-\omega_0)^2} \frac{e^{-\frac{t_c}{t_{ca}}}}{t_{ca}} dt_c$$

show that the corresponding line shape function is

$$g(\omega,\omega_0) = \frac{1}{\pi} \frac{t_{ca}}{1+{t_{ca}}^2(\omega-\omega_0)^2}$$

1.29. Given

$$p(u_x)du_x = \left(\frac{m_m}{2\pi k_B T}\right)^{\frac{1}{2}} e^{\left(-\frac{m_m {u_x}^2}{2k_B T}\right)} du_x$$

show that the lineshape function for Doppler broadening is

$$g(\omega,\omega_0) = \frac{c}{\omega_0}\left(\frac{m_m}{2\pi k_B T}\right)^{\frac{1}{2}} e^{\left[-\frac{m_m c^2}{2k_B T}\left(\frac{\omega-\omega_0}{\omega_0}\right)^2\right]}$$

1.30. Show that the full width at half maximum for Doppler broadening is

$$\Delta\omega_0 = 2\Delta\omega = 2\omega_0\left(\frac{2k_B T \ln 2}{m_m c^2}\right)^{\frac{1}{2}}$$

1.31. Determine the number of oscillations that an atom will undergo over its lifetime, τ_{sp}, of 1.5 μs, if the transition frequency is 10^{13} Hz.

1.32. Determine the frequency that corresponds to the transition wavelength ($\lambda = 632.8$ nm) of the helium–neon laser. Further, determine the value of p of the laser mode closest to the line center if the separation between the mirrors of the laser cavity is 60 cm. How many longitudinal modes will be contained in the linewidth of the laser if the gain curve has a linewidth of 1.5 GHz?

1.33. The transition wavelength of a CO_2 molecule at a temperature of 400 K is $\lambda = 10.6$ μm. Determine its Doppler linewidth.

1.34. For the conditions in Problem 1.33, determine the pressure at which the Doppler and collision broadening mechanisms for the CO_2 laser contribute equally to the linewidth if collision broadening is approximately 6 MHz/Torr.

1.35. A CO_2 laser undergoes a transition at $\lambda = 10.6$ μm at a temperature of 400 K and pressure of 15 Torr. Which broadening mechanism will be dominant under these conditions? The CO_2 molecule has a radius of 116.3 pm, and the upper-level spontaneous emission lifetime is 0.4 ms. Repeat the calculation for a pressure of 50 Torr and compare.

1.36. Given

$$E_1(t) = E_0 e^{\left(i\omega_0 - \frac{\omega_0}{2Q_f}\right)t}$$

show that the spectral density of the signal intensity is

$$I(\omega) = |E_1(\omega)|^2 = \frac{1}{4\pi^2} \frac{{E_0}^2}{\left(\frac{\omega_0}{2Q_f}\right)^2 + (\omega - \omega_0)^2}$$

1.37. Knowing that

$$I(t) = {E_0}^2 \frac{\sin^2\left(\frac{M_c \delta\omega t}{2}\right)}{\sin^2\left(\frac{\delta\omega t}{2}\right)}$$

show that the pulse width (full width at half maximum) is given by:

$$\Delta\tau_p \approx \frac{2\pi}{M_c \delta\omega} = \frac{2\pi}{\Delta\omega_0}$$

1.38. (a) Would you expect the Q-value of a cavity to be high or low when a shutter in front of one of the mirrors is closed? Explain.
(b) Why is it that a mode-locked laser is considered to act as a pulse that propagates back and forth in the cavity?

1.39. If a Q-switched Nd:YAG laser (wavelength = 1.06 μm) generates a rectangular pulse with an energy content of 10 J and of duration 20 ps, what will be the peak output power?

1.40. If the laser in Problem 1.39 is to have an output pulse energy of 7.5 J, determine the fraction of the total population that will be needed for the population inversion if the active medium contains 10^{20} Nd^{3+} ions. Assume that each ion in the upper state undergoes only one transition while the pulse is being generated.

1.41. If an Nd:YAG laser has 500 longitudinal modes in a cavity of length 1500 mm, what will be the width of pulses generated by mode locking? Determine the time-spacing between the pulses.

1.42. For the Nd:YAG laser in Problem 1.41, determine the cavity Q-value if the reflectances of the resonator mirrors are 100% and 80%.

1.43. If 200 of the axial modes in the laser in Problem 1.42 are mode locked, determine the spacing between pulses, time between pulses, and frequency bandwidth.

1.44. Given that the optical path difference between two rays is

$$l = l_2 - l_1 = \sqrt{{x_2}^2 + (z + z_b)^2} - \sqrt{{x_2}^2 + (z - z_b)^2}$$

show that for $x_2 \gg z, z_b$,

$$l \approx \frac{2zz_b}{x_2}$$

1.45. Show that the critical distance, d_c, at which the interference fringes of two narrow slits that are close to each other will nullify each other is given by

$$d_c = \frac{x_1 \lambda}{4 z_b}$$

1.46. (a) Show that for a frequency bandwidth of
$\Delta\nu = 500\,\text{Hz}$
the corresponding wavelength bandwidth is
$\Delta\lambda = 0.6 \times 10^{-12}\,\mu\text{m}$ at $\lambda = 0.6\,\mu\text{m}$

(b) Determine the coherence length of a laser which has a transition wavelength of $\lambda = 0.63\,\mu\text{m}$ and a bandwidth of 1.5 nm. If a coherence length of 2.5 m is desired, what should the bandwidth be?

1.47. Given that the beam size from a confocal resonator is given by:

$$w(x) = w_w \left[1 + \left(\frac{2x}{L}\right)^2\right]^{1/2}$$

and

$$w_w = \left(\frac{L\lambda}{2\pi}\right)^{1/2}$$

show that the beam divergence may be approximated by:

$$\theta_d \approx \frac{\lambda}{\pi w_w}$$

1.48. (a) Given the fact that the bandwidth of a good He–Ne laser is about 500 Hz, while that of an ordinary light source is about 10 GHz, explain why one would result in good interference fringes while the other would not.

(b) What will be the coherence time and length of an Nd:YAG laser of linewidth 195 GHz?

1.49. (a) Determine the divergence of an Nd:YAG laser beam of wavelength 1.06 μm and diameter 5 mm if its divergence is equivalent to that of a plane monochromatic wave that is incident on a circular aperture of the same diameter.

(b) Determine the diameter that a diffraction-limited red He–Ne laser beam will have on Venus, which is at a distance of about 48×10^6 km from the earth, after going through the 500 mm aperture of a telescope.

1.50. What is the distance between the slits B_1 and B_2 if screen B is located at a distance of 1500 mm from screen C in Young's experiment, and the interference fringes observed have a period of 2 mm for a helium-neon laser of wavelength 0.633 μm?

1.51. Make a list of applications (one in manufacturing and one general) for which the following laser properties are useful:

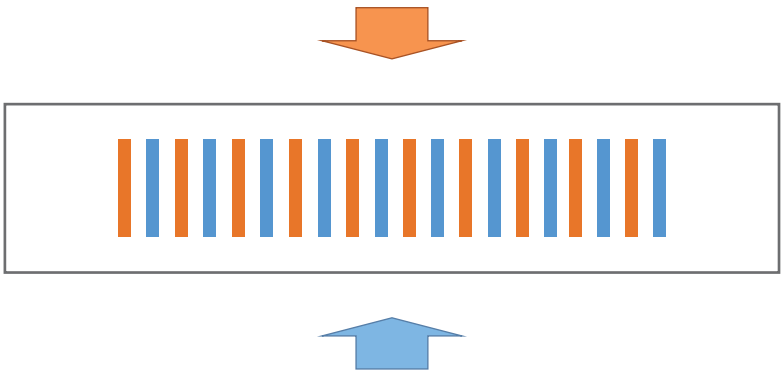
(a) Spatial coherence.

(b) Temporal coherence.

(c) Monochromaticity.
(d) Divergence.
(e) Gaussian mode.
(f) Multimode.

1.52. A modern artist decides to use CO_2 and Nd:YAG laser beams in one of his art displays, by producing the pattern shown below. Let the blue color represent the CO_2 laser and the orange color, the Nd:YAG laser. The spacing between contiguous fringes is constant at 25 mm for both lasers. In using Young's experimental setup, the arrangement is such that the distance of the slits from the x-axis (z_b) can be independently altered for both lasers. However, the distance between the frame containing two slits and the interference pattern in the x-direction (x_2) can only be altered equally and simultaneously for both lasers. Assuming the total room length available for the experimental setup is 50 m,

(a) Find the ranges for z_{b-CO_2} and $z_{b-Nd:YAG}$ if minimum $x_2 = 10$ m for esthetic appeal.
(b) If esthetics were ignored and the full room was to be occupied, then find the relationship between z_{b-CO_2} and $z_{b-Nd:YAG}$.
(c) Plot the relationship between z_{b-CO_2} and $z_{b-Nd:YAG}$ for part (b).
(d) Explain what other problems are encountered in part (b) other than esthetics.
(Ignore the x_1 distance in both cases for setting up the experiment).



BIBLIOGRAPHY

Barrett, C.R., Nix, W.D., and Tetelman, A.S. (1973). *The Principles of Engineering Materials*. Englewood Cliffs, NJ: Prentice Hall.

Berlien, H.-P. and Muller, G.J. (2003). *Applied Laser Medicine*. Berlin: Springer-Verlag.

Chryssolouris, G. (1991). *Laser Machining: Theory and Practice*. Berlin: Springer-Verlag.

Goeppert-Mayer, M. (1931). *Annalen der Physik* 9: 273.

Henderson, R.A. (1997). *A Guide to Laser Safety*. London, UK: Chapman & Hall.

Laser Institute of America. (2007). *American National Standard for Safe Use of Lasers*, ANSI Z136.1-2007. Orlando, FL: Laser Institute of America.

Laud, B.B. (1985). *Lasers and Non-Linear Optics*. New Delhi: Wiley Eastern Limited.

Luxon, J.T. and Parker, D.E. (1985). *Industrial Lasers and Their Applications*. Englewood Cliffs, NJ: Prentice Hall.

O'Shea, D.C., Callen, W.R., and Rhodes, W.T. (1977). *Introduction to Lasers and Their Applications*. Reading, MA: Addison-Wesley.

Shimoda, K. (1986). *Introduction to Laser Physics*, 2e. Berlin: Springer-Verlag.

Svelto, O. (2010). *Principles of Lasers*, 5e. LLC, New York: Springer Science+Business Media.

Thyagarajan, K. and Ghatak, A.K. (2010). *Lasers, Theory and Applications*, 2e. LLC, New York: Springer Science+Business Media.

Wilson, J. and Hawkes, J.F.B. (1987). *Lasers - Principles and Applications*. New York: Prentice Hall.

2 Types of Lasers

There are various types of lasers. Each one has different characteristics, which depend on the active medium used for laser action. In this chapter, we first discuss the major types of lasers, based primarily on the active medium. In addition, we also discuss recent developments in laser technology, especially with regard to industrial lasers, specifically those used for materials processing. The principal laser categories include the following:

1. Solid-state lasers.
2. Gas lasers.
3. Semiconductor (diode) lasers.
4. Liquid dye lasers.

In the following sections, we shall provide highlights on the solid-state, gas, and semiconductor lasers, since they are more widely used for industrial applications.

Even though semiconductor lasers are also based on solid material, they are considered separately since the pumping and lasing mechanisms are different. In each category, a number of lasers have been developed using different active media. The different laser types will be discussed in this chapter in the order listed, in the following sections. The basic concepts related to each general type will be discussed using one or two examples for illustration. The basic characteristics of some major industrial lasers are summarized in Appendix 2.B, and the wavelength ranges of common lasers are summarized in Appendix 2.C.

The notations used to represent the electron configurations in other laser texts are usually complex. In this book, we use an unconventional but very simple representation, which starts with the letter E with a subscript that designates the energy level, for example, E_x, indicating energy level x.

In addition to the major laser categories, we shall discuss new developments in industrial laser technology. Ultrafast (utrashort pulse or femtosecond), slab, disk, and fiber lasers are considered separately under this category. Ultrafast lasers (which are also based on solid-state laser technology) use special techniques, in addition to mode locking, to achieve the ultrashort pulse duration. Slab and disk lasers are also given special consideration, even though they both use solid-state active media. Slab lasers can also use gas active media and may be optically or electrically pumped. Finally, fiber lasers are also based on solid-state laser technology and can generate ultrashort pulse outputs.

2.1 SOLID-STATE LASERS

Solid-state lasers normally use an insulating crystal or glass as the host lattice. The active medium is embedded in the host material as either a dopant or impurity. The dopant is normally a transition

Principles of Laser Materials Processing: Developments and Applications, Second Edition. Elijah Kannatey-Asibu, Jr.
© 2023 John Wiley & Sons, Inc. Published 2023 by John Wiley & Sons, Inc.
Companion Website: www.wiley.com/go/Kannatey-Asibu/PrinicplesLaserMaterialsProcessing

metal or rare-earth element and substitutes for some of the atoms in the host material, rather than being an interstitial impurity. For the solid-state lasers that we consider under this section, the active medium is shaped either in the form of a rod or a slab. We use the rod form as illustration in most of our discussion. The slab configuration will be discussed at the end of this section. The rod is normally cylindrical in shape, with ends which are ground and polished to be plane and parallel. Its dimensions depend on the active medium being used. The ends of the rod may either be silvered (one completely and the other partially), in which case the rod constitutes the optical cavity; or the rod may be placed between two external mirrors. Pumping is commonly done using a flashlamp, and the configuration involves either a cylindrical cavity with elliptical cross section, or a helical flashtube. The rod temperature may be controlled by circulating air or liquid around it. Otherwise, the heat generated can change the cavity dimensions and consequently, the cavity modes.

Most solid-state lasers generate pulsed beams, even though some generate continuous-wave (CW) outputs. Their coherence lengths are thus relatively short, making them unsuitable for a number of interference-based applications. However, the significant amount of energy available in each pulse makes them attractive to applications that require such energy bursts. These include resistor trimming, initiation of thermonuclear fusion, and spectroscopic research. Common types of solid-state lasers include the Nd:YAG and Nd:Glass lasers. These are further discussed in the following subsections. The ruby laser was the first laser to be successfully made and operated. However, since it is not extensively used anymore, it is not further discussed.

Neodymium (Nd) lasers have either a crystal or glass material as the host lattice, and this is doped with neodymium ions, Nd^{3+}, that constitute the active medium. The energy levels are thus those of the neodymium ions, as illustrated in Figure 2.1. These are four-level lasers. The ground level (level 0) is the E_0 energy level. Again, the pump level (level 3) has two bands, with E_{3a} having a mean wavelength of 0.73 μm, while E_{3b} has a mean wavelength of 0.8 μm. The upper lasing level (level 2) is E_2 with a lifetime of about 0.23 ms, and from there, transition occurs to the lower lasing level (level 1), E_1. This transition is primarily homogeneously broadened with a full width at half maximum (FWHM) of about 195 GHz, and a transition wavelength of 1.06 μm in the infrared range. Atoms are pumped from level 0 to level 3, from which they undergo rapid non-radiative decay to level 2 which is a metastable state. The energy difference between level 1 and the ground state, level 0, is such that $h_p\nu \gg k_B T$ at room temperature, and thus, transition from level 1 to level 0 is very fast, occurring by non-radiative processes so that level 1 is practically empty under thermal equilibrium conditions. Any atoms that make a transition to level 2 then essentially result in a population inversion. There is a variety of host materials that are used for neodymium lasers, but the most common ones are yttrium aluminum garnet (YAG) and glass.

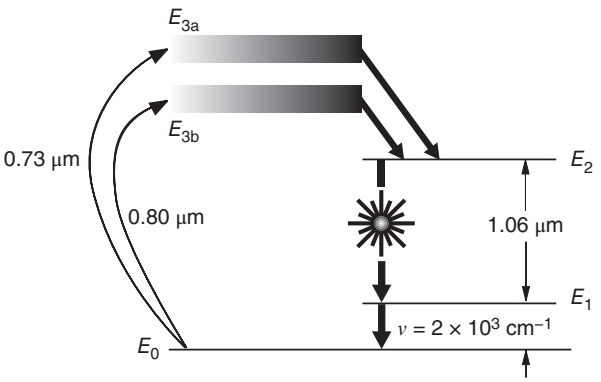


Figure 2.1 Energy levels of neodymium lasers.

2.1.1 The Nd:YAG Laser

The host lattice for this laser is a crystal of YAG with the chemical composition $Y_3Al_5O_{12}$, where the Nd^{3+} ions (about 0.1 to 2%) substitute for some of the Y^{3+} ions. Operation of the Nd:YAG laser can be either in the CW or pulsed mode, depending on whether pumping is continuous or intermittent. A xenon flashlamp may be used at medium pressure (500–1500 Torr) or a krypton lamp at high pressure (4–6 atm), and the rod size ranges from 0.5 to 1.0 cm in diameter and from 5 to 20 cm in length. The power output in CW mode varies from 150 W to 6 kW, and that in Q-switched pulsed mode is of the order of 50 MW with a pulse duration of about 20 ps, at a repetition rate of 1–100 Hz. The high-power (up to 2 kW) CW laser may be achieved by having three Nd:YAG rods in line as a single oscillator, with each rod being arc lamp-pumped. The very high powers (up to 6 kW) are obtained by pumping with a diode laser, Figures 1.23 and 1.24. The efficiency of both pulsed and CW Nd:YAG lasers typically ranges between 1 and 3%. Applications of Nd:YAG lasers include laser surgery and materials processing, for example, welding, cutting, drilling, and surface modification.

2.1.2 The Nd:Glass Laser

The host lattice for the Nd:Glass laser is glass such as silicate, phosphate, or fluoride glass, with 1–5% Nd^{3+}. It has the advantage that the rod size can be much larger (up to 1 m in length and about 50 mm in diameter) than that for Nd:YAG, since glass can be easily made with high quality or optical homogeneity (free of residual stresses) to large sizes because of its lower melting temperature. With power output per unit volume being comparable to that of Nd:YAG, relatively larger power outputs can be obtained with this type of laser. Since glass has an amorphous structure, it has short range order. Thus, the environment and therefore the field of each ion are different from that of any other ion. This variation of ion environments in the glass matrix results in additional inhomogeneous broadening, and thus a much broader bandwidth. The Nd:glass laser therefore normally operates in multimode. Very short pulse periods, of the order of 5 ps, with high output powers, can thus be obtained when the output is mode-locked.

One major disadvantage of Nd:glass lasers is the low thermal conductivity of glass, which is about an order of magnitude smaller than that of Nd:YAG. This limits their usefulness for CW operation, and are thus only used for low-rate-pulsed outputs, with a typical output pulse rate of 1 Hz. The relatively high peak pulse outputs (up to 20 TW) of Nd:glass lasers enable them to be used in laser-induced nuclear fusion reactions. However, the low pulse rates limit their applicability in manufacturing to such operations as drilling and spot welding, and in the cases where they are acceptable, are able to produce superior quality holes at depth-to-diameter ratios up to 50:1, much higher than what can be achieved with the Nd:YAG laser.

2.2 GAS LASERS

Gas lasers are among the most common form in the laser industry. The power levels range from several kilowatts (carbon dioxide (CO_2) lasers) to milliwatts (helium-neon (He-Ne) lasers). They can be operated either in the continuous mode or pulsed mode, with output frequencies ranging from ultraviolet to infrared.

As the name implies, gas lasers use a gaseous material as the active medium. The broadening mechanisms in gas lasers are not as strong as those in solids. Thus, the resulting linewidths, determined primarily by Doppler broadening, are relatively small. This is because collision broadening is relatively small due to the low pressures normally used in gas lasers. As a result, the energy levels are relatively narrow, and thus a sharp emission line is essential for excitation. Optical pumping, with its broad emission spectrum, is therefore not suitable for pumping gas lasers, since it would

result in inefficient pumping. Electrical pumping is thus the most common means of exciting the active medium in gas lasers. Pumping is also done by chemical means, with an electron beam, or by gas-dynamic expansion.

When classified by the active medium, one can identify four principal types of gas lasers:

1. Neutral atom lasers.
2. Ion lasers.
3. Metal vapor lasers.
4. Molecular lasers.

2.2.1 Neutral Atom Lasers

The most common example of neutral atom lasers (using inert gases as the active medium) is the He-Ne laser, which was also the first laser to generate CW output. The active medium in this case consists of one part neon to 10 parts helium. This is a four-level laser, and Figure 2.2 illustrates the energy level schemes of He and Ne. The helium atoms are more easily or efficiently excited to the higher levels by electron collision than are neon atoms. Thus, the electrons that are accelerated by the passage of a discharge through the mixture excite the helium atoms to the metastable higher energy levels denoted by HE_{3a} and $HE_{3b.}$ These energy levels of helium happen to coincide with some of the excited states of neon, the NE_{3a} and NE_{3b} levels. The excited helium atoms are thus able to transfer energy to neon atoms which are in the ground state when collision takes place between the two. This results in de-excitation of the helium atoms involved to the ground state, while the neon atoms are excited to the higher energy level. This process of energy transfer through collision is referred to as *resonant collision* or *resonant energy transfer* and is so called because the corresponding energy levels coincide. The process increases the population of the NE_{3a} and NE_{3b} levels of neon relative to the lower NE_{2a} and NE_{2b} levels, resulting in a population inversion between the two sets of levels.

Lasing action occurs between energy levels of the neon atoms. The role of helium is thus to assist in the pumping process. The metastable state of the excited helium atoms (staying at the high level for a relatively long time) makes the energy transfer process more efficient. The decay time of the

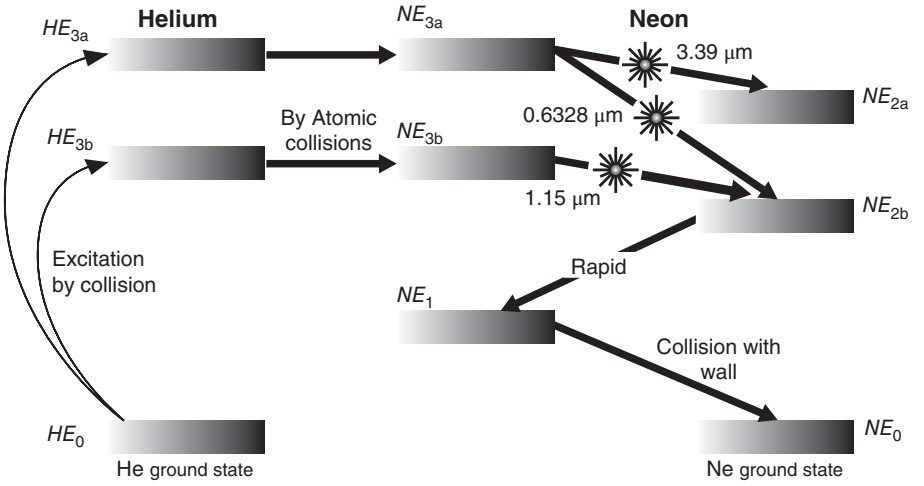


Figure 2.2 Energy levels of the He-Ne system.

upper lasing levels of neon (about 100 ns) is about an order of magnitude greater than that of the lower lasing levels (about 10 ns), thereby satisfying a necessary condition for CW operation.

The principal or strongest transitions are shown in Figure 2.2, with transition at wavelengths of 0.633, 1.15, and 3.39 μm. The last two wavelengths fall in the infrared region, while the first one results in the red light and is the most common mode in which the laser is used. The desired wavelengths can be selected using techniques discussed in Section 1.2.2.

After transition, the atoms in the lower lasing level relax spontaneously to the NE_1 level which is metastable. Thus, if the temperature of the neon gas is high enough, electrons from the NE_1 state may jump back to the NE_{2a} and NE_{2b} levels (Boltzman's law). This phenomenon is known as radiation trapping and will cause the population inversion between the upper lasing level and the NE_{2a} and NE_{2b} levels to decrease, resulting in quenching.

The three main factors which affect radiation trapping and the population inversion of He-Ne lasers are as follows: current density; total and partial gas pressures of He and Ne; and diameter of the discharge tube. With a higher current density in the discharge tube, the temperature of the gas becomes higher. Therefore, there is an optimum current density that provides maximum pumping power without quenching laser action.

The gas pressures of He and Ne influence the population of the NE_1 state. If the population of the NE_1 state is increased due to a pressure change, the chances of radiation trapping are increased. The optimum pressures have been found to be about 1 mm Hg (1 Torr) for helium and about 0.1 mm Hg for neon. Finally, in order to reduce the population of the NE_1 level and thus reduce the chances for radiation trapping, the NE_1 neon atoms must collide with the discharge tube walls to de-energize. If the inner diameter of the discharge tube wall is too large, there will be less chance of collisions with the wall. Thus, the tube wall must be made small enough to ensure a high-enough probability of wall collisions. In practice, tube diameters range from about 1–6 mm. Beyond that range, the output power is reduced, because the population inversion is lessened due to a fewer number of wall collisions by atoms in the NE_1 state.

Figure 2.3 is a schematic of the He-Ne laser. The discharge tube is typically 20–80 cm in length. At the ends are the mirrors, which may be either internal (sealed inside the discharge tube) or external to the sealed tube (Figure 2.3). The internal system has the disadvantage that it needs to be replaced periodically due to erosion by the discharge. On the other hand, with the external system, the windows at the ends of the tube reflect part of the beam away, resulting in losses. This problem is mitigated by positioning the windows at the Brewster angle, θ_B (Section 3.3), which is given by:

$$\tan \theta_B = n \tag{2.1}$$

where n is the refractive index of the window material.

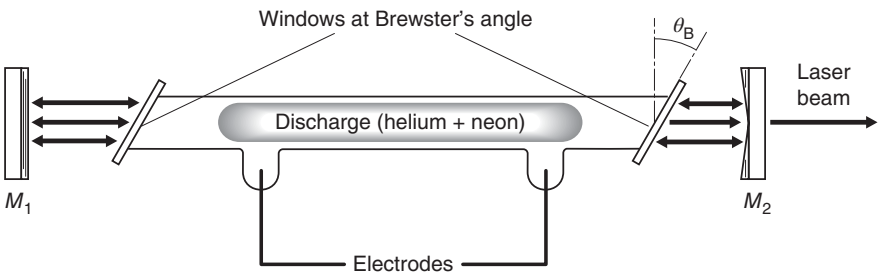


Figure 2.3 Schematic of a He-Ne laser (external mirrors).

The output beam of the He-Ne laser has a bandwidth of about 1.6 GHz due primarily to Doppler broadening. Thus, there may be several axial modes operating simultaneously. A single-axial-mode operation can be obtained by reducing the cavity length to about 10–15 cm.

The outputs of He-Ne lasers typically range from 0.5 to 50 mW, with an input power requirement of 5–10 W. The overall efficiency is thus of the order of 0.02%.

Common applications for He-Ne lasers include position sensing, character or bar code reading, alignment, displacement measurement by interferometry, holography, and video disk memories.

2.2.2 Ion Lasers

Ion gas lasers are generally four-level lasers, and the active medium is an ionized inert gas, with a typical operating pressure of about 1 Torr. There are a variety of ion lasers, including argon, krypton, xenon, and mercury ion lasers. These have essentially the same design. We shall use the argon laser for illustration. The energy level scheme of the argon ion (Ar^+) laser is shown in Figure 2.4. Excitation to the higher laser level, E_3, can occur in three ways:

1. Through collision of (Ar^+) ions in their ground state with electrons.
2. Through collision of (Ar^+) ions in the metastable state with electrons.
3. Through radiative cascading from higher levels (to the E_3 level).

Laser transition occurs between the E_3 (higher) and E_2 (lower) levels, with the lifetime of the E_3 level being about 10^{-8} s which is an order of magnitude greater than that of the E_2 level (about 10^{-9} s), thereby satisfying the condition for CW operation.

Unfortunately, the pumping process involves two stages instead of one. In the first stage, the neutral atom is ionized by colliding with an electron in the discharge tube. The ion may then be excited to a higher energy level by another collision with an electron, in the second stage. Since pumping requires two stages, the pumping process, and therefore the ion laser in general, is inefficient. The efficiency is approximately 0.1%. Current densities of around 1000 A/cm^2 are therefore required to maintain the threshold population inversion in ion lasers.

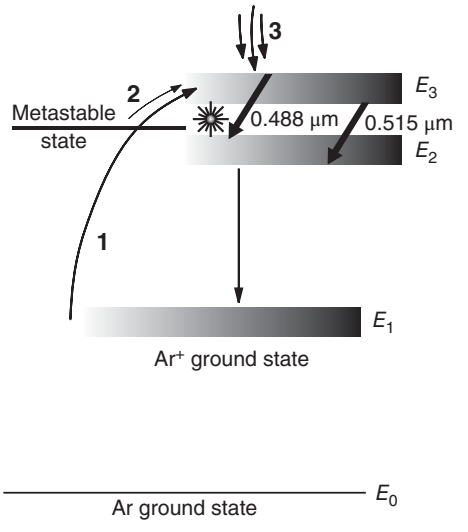


Figure 2.4 Energy levels of the Ar^+ system.

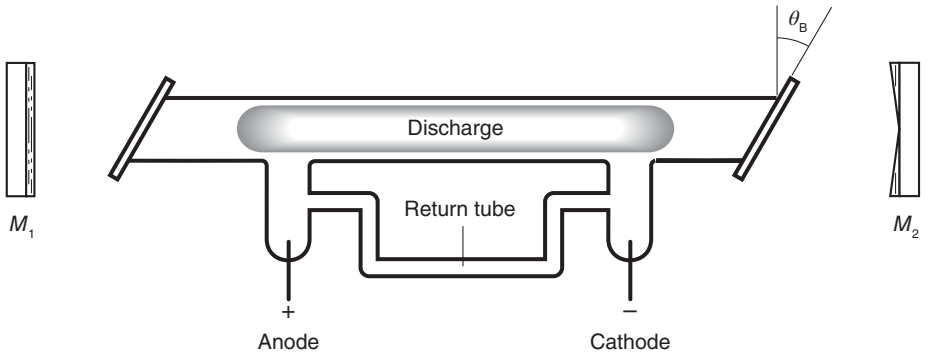


Figure 2.5 Schematic of an Ar^+ laser.

In addition to raising the temperature of the discharge tube, the high current density causes the ions to migrate toward the cathode. Since the ions have a lower mobility than electrons, there is a tendency for the ions to accumulate at the cathode. This might result in the discharge being extinguished due to a loss of active medium. To prevent this from happening, a return path is provided between the anode and the cathode (Figure 2.5). The return tube is made longer than the discharge tube to prevent the discharge from passing through the return tube (path of least resistance). Overall, the two-stage pumping scheme of ion lasers significantly complicates their design and restricts their performance.

Output of the Ar+ ion laser occurs at a number of wavelengths, with the strongest being the 0.4881 μm (blue) and 0.515 μm (green) wavelengths. Power outputs typically range from about 1 to over 20 W CW but can also be lower than 1 W. Ar+ ion lasers are used in laser printers, in surgery, for pumping dye lasers, and in spectroscopy. Other ion lasers include the krypton (Kr+) ion laser with an output wavelength of 0.6471 μm and a CW output of 5–6 mW, and the xenon (Xe^+) ion laser with an output wavelength of 0.5395–0.995 μm, and a pulsed peak power output of about 200 W.

2.2.3 Molecular Gas Lasers

Unlike the neutral atom and ion gas lasers, molecular gas lasers derive their transitional energy from a molecule. The overall energy of a molecule is composed of four principal components:

1. Electronic energy, which results from electron motion about each nucleus.
2. Vibrational energy, which results from vibration of the constituent atoms about an equilibrium position.
3. Rotational energy, which results from rotation of the molecule about an axis.
4. Translational energy, which results from linear or curvilinear motion of the molecule.

Each electronic energy level (Figure 2.6a) consists of a number of vibrational levels (Figure 2.6b), and each vibrational level in turn consists of a number of rotational levels (Figure 2.6c). In Figure 2.6, E_1 refers to the ground level, and E_2 refers to the first excited state for the case where the atoms of the molecule are held fixed at a nuclear separation distance. The electronic, vibrational, and rotational energies are quantized, and transitions between the various levels can produce laser oscillations. Translational energy, on the other hand, is not quantized and is thus not useful for laser oscillations.

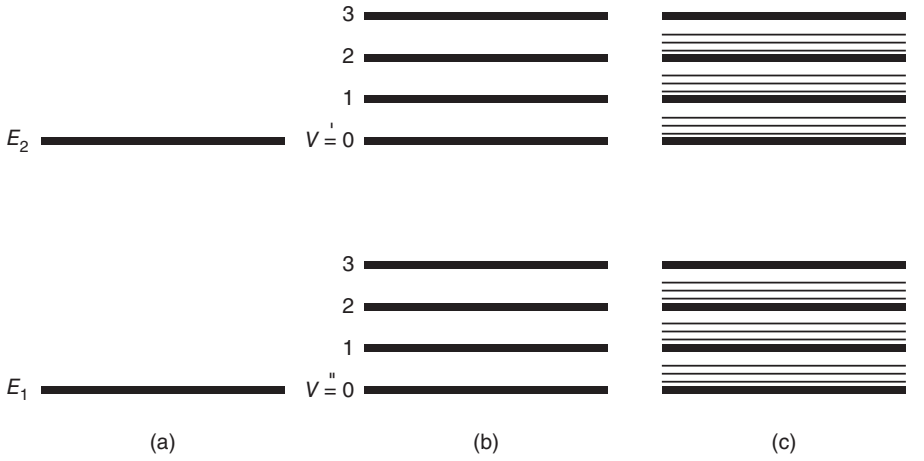


Figure 2.6 Energy levels of a molecule. (a) Electronic. (b) Vibrational. (c) Rotational.

There are three common transitions associated with the quantized energies. These are normally used to categorize the lasers as:

1. Vibrational-rotational.
2. Vibronic (vibrational-electronic).
3. Pure rotational.

A fourth type of molecular laser, which falls in a category of its own, is the excimer laser. In the following sections, we shall discuss the vibrational-rotational and excimer lasers, since these are more easily produced commercially. Since laser action is relatively difficult to achieve in rotational lasers, these are not further discussed.

2.2.3.1 Vibrational-Rotational Lasers In vibrational-rotational lasers, oscillations result from transitions between vibrational levels (rotational level of one vibrational state and the rotational level of another vibrational state) of a given electronic state, normally the ground state. The wavelength of the output beam ranges from 5 to 300 μm (middle to far-infrared).

The CO_2 laser is the most common example. It has the advantage of relatively high efficiencies, about 10–30%. Common applications of the CO_2 laser include materials processing, communications, spectroscopy, and surgery.

Figure 2.7 shows the following three fundamental vibration modes of the CO_2 molecule:

1. Asymmetric stretching.
2. Bending.
3. Symmetric stretching.

The energy level or oscillation behavior of the system is described using the number of energy quanta in each vibrational mode. For example, (*pqr*) indicates that there are *p* quanta in the symmetric mode, *q* quanta in the bending mode, and *r* quanta in the asymmetric mode.

The active medium for the CO_2 laser is a mixture of CO_2, N_2, and He. The CO_2 molecule provides the transitions that generate the laser beam. N_2 helps to increase the population of the upper lasing

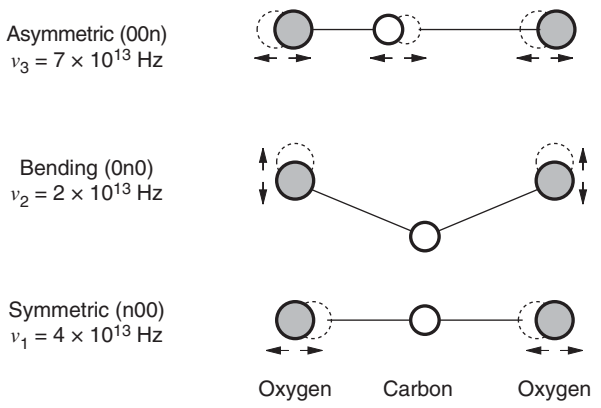


Figure 2.7 Vibrational modes of the CO_2 molecule.

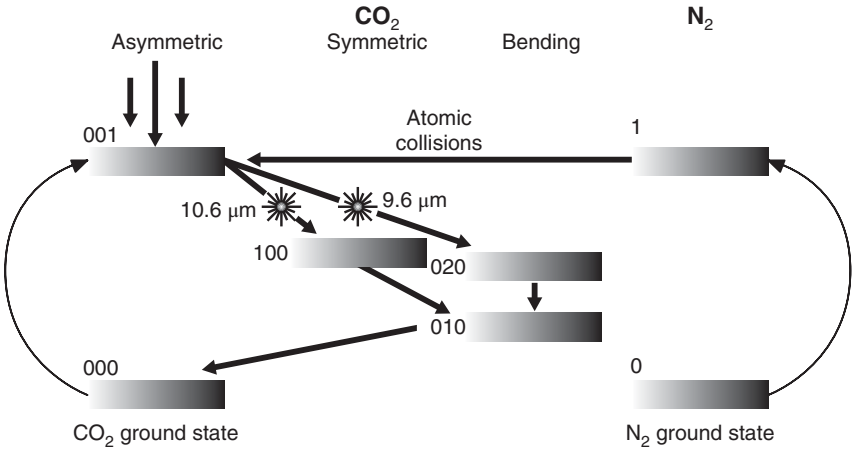


Figure 2.8 Energy level scheme of the CO_2 laser.

level, while He helps to depopulate the lower lasing level, thereby enhancing the achievement of population inversion. The contributions of the individual gases are now further discussed.

The energy level scheme of the electronic ground state for this type of laser is shown in Figure 2.8. The (001) level of the CO_2 molecule, which is the upper lasing level, is pumped by the following:

1. Direct electron collision.
2. Resonant energy transfer.
3. Radiationless transition from upper excited levels.

Collision between electrons accelerated in the discharge and the CO_2 molecules result in the molecules being excited preferentially to the (001) energy level. This process can be represented as:

$$e^- + CO_2(000) \longrightarrow e^- + CO_2(001) \tag{2.2}$$

With resonant energy transfer, electron collision with an N_2 molecule first results in the latter being excited from the ground level (0) to the upper level (1). This is a very efficient process, and

since the upper level is metastable and has about the same energy level as the (001) level of CO_2, resonant transfer of energy can occur easily between the (1) level of N_2 and the (001) level of the CO_2 molecule. Finally, higher excited levels (00r) of CO_2 decay rapidly to the metastable (001) state. As a result of these various processes, pumping of the (001) level is very efficient, contributing to the relatively high efficiency (up to 30%) of the CO_2 laser.

The upper lasing level is (001), and the lower levels are (100) and (020). Transition between the (001) and (100) levels results in oscillation at the wavelength 10.6 μm, while that between the (001) and (020) results in oscillation at the wavelength 9.6 μm, with the transition at 10.6 μm having a much greater likelihood (about 20 times) of occurring than the 9.6 μm transition. The 10.6 μm output is thus more common. Frequency selection techniques can be used to suppress the 10.6 μm output if the 9.6 μm output is desired.

In reality, each of the vibrational levels mentioned, say (001), is composed of several rotational levels which have a population distribution given by the Boltzmann distribution. Only the rotational level with highest gain oscillates, since the thermalization rate of the rotational levels (about 10^7 s^{-1} $(mm\,Hg)^{-1}$) is much greater than the rate at which the population of the oscillating rotational level decreases. Thermalization refers to redistribution of the population of various levels to achieve equilibrium or steady-state distribution.

The lifetimes of the various levels are influenced by the partial pressures of the gas mixture. For a typical mixture with partial pressures of 1.5 mm Hg for CO_2, 1.5 mm Hg for N_2, and 12 mm Hg for He, the lifetime of the upper lasing level is about 0.4 ms. The energy difference between the lower lasing levels, (100) and (020), is very small, much less than $k_B T$, and thus the (100) → (020) decay is very fast. Likewise, the (100) → (010) and (020) → (010) decays are very fast. Thermal equilibrium is thus reached between the three levels (100), (020), and (010) in a relatively short period. As a result, if the decay from the (010) level to the ground state (000) is not fast enough, then there would be an accumulation of atoms in all three lower levels since they are in thermal equilibrium. This would lead to a quenching of population inversion. The decay of the (010) level to the ground state is accelerated or enhanced by the presence of He in the gas mixture. For the partial pressures mentioned, the corresponding lifetime of the (010), (100), and (020) levels is about 20 μs, therefore satisfying the condition for laser oscillation.

The presence of He in the mixture serves yet another purpose. To prevent repopulation of the (010) level and thus lower lasing levels by thermal excitation, the energy difference between (010) and the ground state has to be much greater than $k_B T$. This occurs if the temperature of the CO_2 gas is kept relatively low. Since He has a high thermal conductivity, it can conduct heat away to the tube walls effectively, thereby keeping the CO_2 gas at a low temperature.

There are various designs for CO_2 lasers, including the following:

1. Sealed tube lasers.
2. Axial (longitudinal) flow lasers.
3. Transverse flow lasers.
4. Transversely excited atmospheric (TEA) lasers.
5. Gas-dynamic lasers.

Others are diffusion-cooled lasers and capillary waveguide lasers. We shall only discuss the first three designs since they are more commonly used as industrial lasers.

Sealed Tube Lasers In sealed tube lasers, the gas mixture or active medium is contained in an enclosed chamber (the discharge tube). Therefore, there is no fresh supply of gases or active medium

to the laser. In the case of CO_2 lasers, dissociation products that are undesirable for laser operation may form in the tube. For example, carbon monoxide (CO) forms from CO_2:

$$2CO_2 \longrightarrow 2CO + O_2 \tag{2.3}$$

The formation of carbon monoxide is undesirable in CO_2 lasers, because it absorbs light at 10.6 μm. Thus, if a large amount of dissociation products exist within the discharge tube, the laser may be quenched. To prevent quenching, a catalyst is used to regenerate CO_2 from CO. There are a number of ways this can be done:

1. A small amount (about 1%) of water vapor (H_2O) is added to the gas mixture.
2. A Ni cathode that is heated to about 300°C is used as the catalyst.

With these catalysts, the laser can be used for over 10,000 hours. Typical output powers for sealed tube lasers are about 60 W/m. The output beam of the sealed tube CO_2 laser normally operates in the low-order mode. Therefore, it has been used in microsurgery and micromachining where the high accuracy of the low-order mode beam is essential. Machining applications include drilling and cutting of thin metal sheets (about 0.5–1.5 mm in steels) and nonmetals.

Axial Flow Lasers Figure 2.9 shows the configuration of a typical axial flow CO_2 laser, where the gas mixture is made to flow continuously through the tube. Continuous flow of the gas mixture is used to enable the dissociation products of the discharge, especially CO, to be removed, thereby preventing contamination. The cavity mirrors can be either internal or external. For the external system, the windows at each end are configured at the Brewster angle. The tube walls are water-cooled to remove heat that is dissipated in the discharge. The pressure of the gas in the tube is kept around 15 mm Hg. Broadening of the output laser is then primarily due to Doppler broadening, with a linewidth of about 50 MHz, even though collision broadening is also significant. Because of the relatively small linewidth, the system can easily be operated in a single longitudinal mode by confining the cavity length to less than 1 m. For mechanical and thermal stability, the tube is mounted in an invar rod frame to reduce expansion during operation. There are two types of axial flow lasers:

1. Slow axial flow lasers.
2. Fast axial flow lasers.

SLOW AXIAL FLOW LASERS In this type of laser, the gas mixture flows at a relatively slow speed (about 20 L/min), with the prime objective of removing the dissociation products. Heat is removed

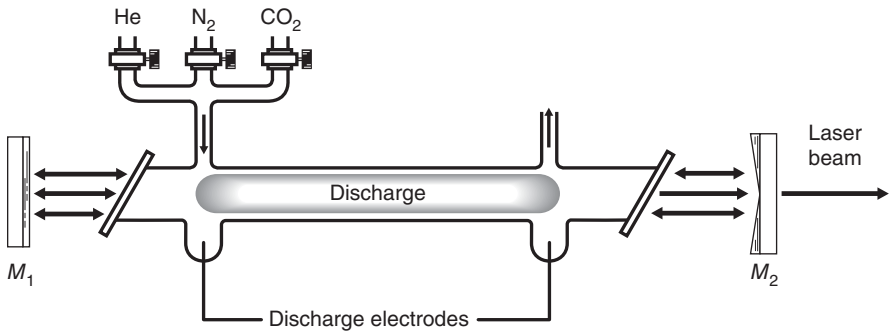


Figure 2.9 Schematic of the axial flow CO_2 laser.

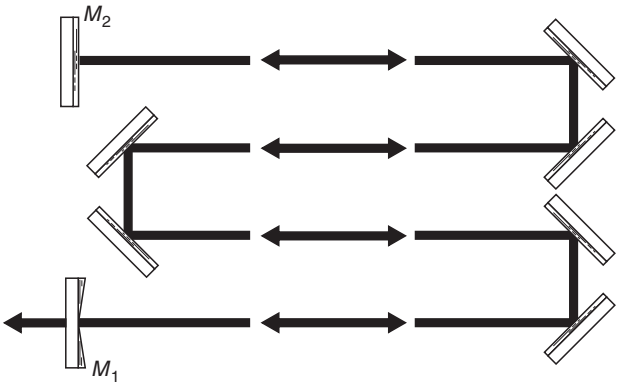


Figure 2.10 One possible resonator configuration.

by conduction from the tube center to the walls. The output power is typically about 50–60 W/m, or 50–500 W overall, and does not depend on the tube diameter, which is about 25 mm. The higher output powers are obtained by increasing the optical path length of the beam in the resonator, using folding mirrors to redirect the beam along different paths (Figure 2.10). The lower power levels are used in laser surgery, while the higher powers are more appropriate for scribing, resistor trimming, welding of thin sheets, and cutting nonmetals.

FAST AXIAL FLOW LASERS The output power of axial flow lasers is limited by the heat generation within the laser, and the manner in which it is extracted. Since the efficiency of the laser is up to 30%, over 70% of the energy input is dissipated as heat which, in slow axial flow lasers, is removed by diffusion from the tube center to the walls. Removal of the dissipated heat can be made more efficient by flowing the gas mixture through the tube at much higher speeds of about 300–500 m/s to carry away the heat by convection. The gas is then recycled through a heat exchanger. This enables the power to be greatly increased to about 600 W/m, or 500 W to 6 kW in overall power output, and units as large as 20 kW have been built. This is a *fast axial flow laser*. Axial flow lasers are extensively used in materials processing, for example, welding and cutting. They tend to have better beam characteristics than transverse flow lasers described in the following section.

Transverse Flow Lasers In transverse flow lasers, the gas flow, electrical discharge or current flow, and resonator axis are mutually perpendicular to one another (Figure 2.11). This is because the heat generated within the discharge tube can be more efficiently extracted by flowing the gas at a rate of about 60 m/s in a direction normal to the resonator axis. This results in a reduced operating temperature, and under such circumstances, the laser can be operated at higher pressures. But the output power for CO_2 lasers increases with the pressure. Thus, for a given discharge current density, higher output powers are achieved by increasing the total pressure from 15 to 100 mm Hg. However, the electric field required to maintain the discharge at the higher pressures also increases to very high levels of about 500 kV/m. For standard axial flow lasers of cavity length 1 m or more, the amount of voltage required would be impractical. To circumvent this, the discharge is applied normal to the resonator axis so that the distance between the electrodes is as small as possible, of the order of 10 mm. This is also sometimes referred to as a transversely excited (TE) laser.

There are two principal ways in which the discharge is sustained in transverse flow lasers.

1. The discharge is said to be self-sustained when the gas is ionized by the discharge. To initiate the discharge, a high electric field needs to be applied. However, the electric field necessary

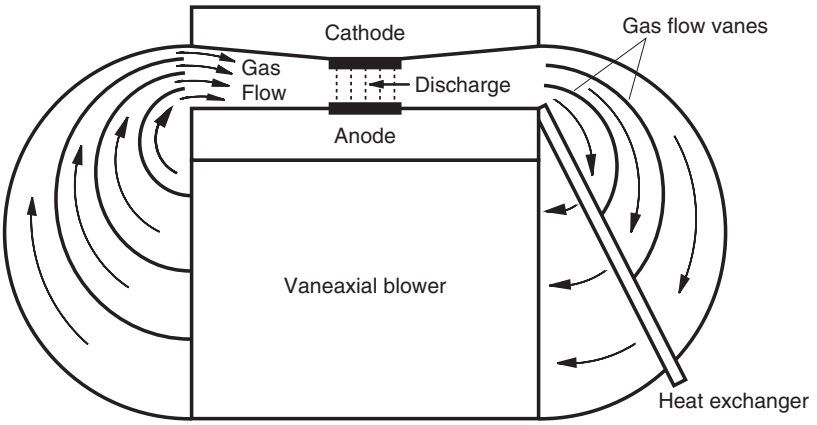


Figure 2.11 Simplified two-dimensional sectional view of a transverse flow CO_2 laser. Source: Svelto (2010)/Springer Nature.

to maintain the discharge is relatively low due to the short distance between the electrodes, but the current is quite high. The output power is about 600 W/m or up to 9 kW overall power. The beam is often folded or reflected back and forth a number of times between the mirrors to increase the power output (Figure 2.10).

2. The non-self-sustained discharge involves the use of an electron beam to ionize the gas, such that the excitation voltage that is applied is just optimum for attaining population inversion, but itself not adequate for sustaining the discharge. Power outputs obtained with this mode of excitation may be about 10 kW/m or up to 15 kW overall power.

The high powers attainable from transverse flow lasers make them highly suitable for materials processing applications such as welding, surface heat treatment, and cladding.

2.2.3.2 Excimer Lasers Excimer lasers are based on the transition of an excited diatomic molecule to a lower energy state where it dissociates into single atoms. The potential energy of a diatomic molecule (dimer), M_2, depends on the separation distance, r, between the nuclei centers of the individual atoms, and is of the form shown in Figure 2.12a. The overall curve is the resultant of repulsive and attractive energy curves. The resultant curve, that is, the potential energy well, has a minimum, Q_0, which occurs at the equilibrium separation distance, r_0, between the atoms. Thus, each electronic energy level has a potential energy well, as shown in Figure 2.12b.

In some molecules, the ground-state potential energy curve does not have a minimum and consists only of the repulsive component, whereas the excited state does have a minimum (Figure 2.12c). Since the ground state does not have an equilibrium separation, the molecule does not exist in this state, as the constituent atoms repel each other in this state. However, the individual atoms, M, may exist in the ground state. On the other hand, since the excited state of the molecule does have a minimum in the energy curve, the molecule can exist in the excited state M_2^*. A molecule that exists under such conditions is referred to as an *excimer* (excited dimer). Common examples are the noble gases (Ar_2^*, Kr_2^*, Xe_2^*) or their oxides, (for example, ArO^*), or halides (for example, ArF^*). The noble gas halide excimer lasers tend to have better beam properties than the pure noble gas excimer lasers.

For such a material, the excimer corresponds to the upper lasing level, while the free atoms represent the lower lasing level. Transition occurs between different electronic states. Thus, when the molecule undergoes a transition from the upper (excimer state) to the lower lasing level

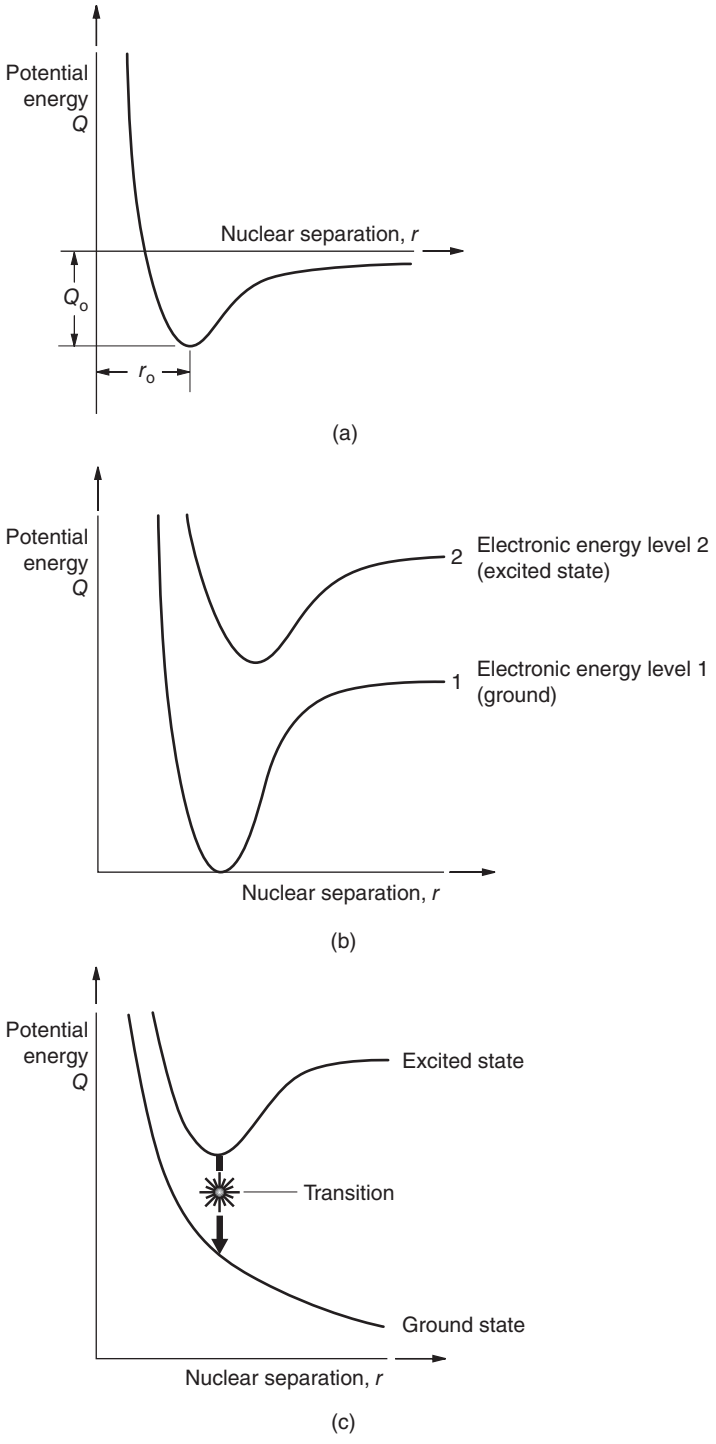


Figure 2.12 (a) Potential energy well of a diatomic molecule. (b) Potential energy for various electronic energy levels. (c) Energy levels of an excimer laser.

(ground state), it dissociates into individual atoms. As a result, the ground state of the molecule is always empty, making it easier to achieve population inversion. The transition is broadband since there are no well-defined vibrational-rotational transitions. The resulting output laser beam is thus tunable over a range of wavelengths.

Pumping may involve subjecting the noble gas to a pressure of about 10 atm and using an electron beam or an electrical discharge to excite it to the higher level, where the atoms combine to form the excited M_2^* excimer. For the Xe_2^* excimer, transition occurs at a wavelength of 0.172 μm with a linewidth of $\Delta\lambda = 0.015$ μm, over which the laser beam is tunable, while the output wavelength of the Ar_2^* excimer laser is about 0.125 μm. The noble gas halide excimer lasers also generate beams in the ultraviolet range. Examples include ArF, 0.193 μm; KrF, 0.248 μm; XeCl, 0.308 μm; and XeF, 0.351 μm. The linewidth of the ArF excimer laser is about $\Delta\lambda = 500$ pm. This can be reduced to about $\Delta\lambda = 1$ pm using wavelength dispersive optical elements in the laser cavity.

The output beam is normally pulsed with a pulse duration of about 10 ns, average output power of up to 200 W, and a pulse rate of up to 1 kHz. The efficiency of the laser is about 1%. Excimer lasers are useful in applications such as isotope separation where the ultraviolet output is very useful. Generally, excimer lasers are useful for removal processes such as photochemical reactions, or more specifically, photoablation, and are typically used for polymers, ceramics, and glass. One application involves making holes in the nozzles of ink jet components for printers. Other applications include wire stripping and burr removal. In recent years, they have also been gaining increasing importance in materials processing.

2.3 SEMICONDUCTOR (DIODE) LASERS

Semiconductor (diode) lasers are based on the generation of photons when electrons in the conduction band of a semiconductor material recombine with holes in its valence band. The discussion in this section starts with a background on semiconductor materials in general. This is followed by the basic principles of semiconductor lasers, and finally the different types of semiconductor lasers.

2.3.1 Semiconductor Background

As discussed previously in Section 1.1.1, electrons in an atom can only occupy specific quantum states, and the distribution of electrons in the atom depends on the temperature. In semiconductors, it is more appropriate to consider the energy levels of the entire solid rather than individual atoms. This is because the available energy levels for electrons in these solids can be grouped into two broadbands, where the energy levels within each band are so closely spaced that they essentially overlap. These are the following:

1. The valence band where the energy levels are almost completely filled by electrons such that the electrons are not freely mobile in the solid.
2. The conduction band where the energy levels are either empty or partially occupied by electrons which are able to move freely within the solid (Figure 2.13).

The energy separation between these two bands is referred to as the bandgap or forbidden band and cannot be occupied by electrons in the solid.

The energy levels in the valence and conduction bands are described by Fermi–Dirac statistics where the probability, $p(E)$, of occupation of an energy level is given by:

$$p(E) = \frac{1}{1 + e^{\left(\frac{E - F_e}{k_B T}\right)}} \tag{2.4}$$

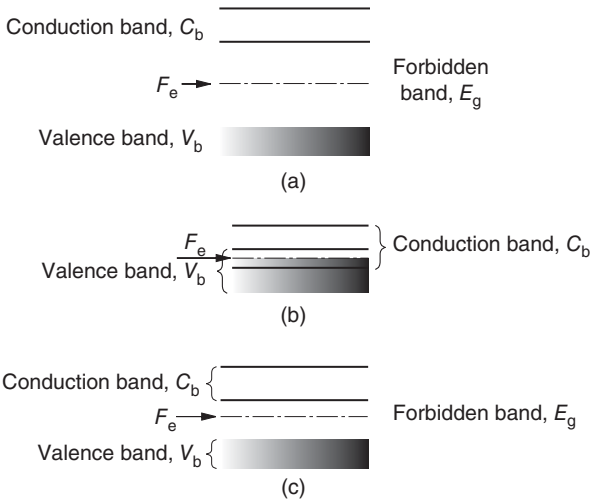


Figure 2.13 Energy bands of: (a) insulators, (b) metals and (c) semiconductors.

where E is the energy level of interest, F_e is energy of the Fermi level, k_B is Boltzmann's constant, and T is absolute temperature.

The Fermi energy defines the energy level where the probability that an electron will be in that state is 0.5. From this equation, it is found that at absolute zero:

$$p(E) = \begin{cases} 0 & \text{for} \quad E > F_e \\ 1 & \text{for} \quad E < F_e \end{cases} \tag{2.5}$$

At absolute zero temperature, the electrons occupy the lowest possible energy levels allowed, with all lower energy levels being completely filled before the next level is occupied. Under these conditions, the valence band is completely occupied while the conduction band is empty, and the boundary between them is the Fermi energy level. The material is then an insulator (Figure 2.13a). This is because conduction of current results from flow of electrons from one level in a band to another level in the band, when the material is subjected to an electric field. With the valence band being completely occupied, then the electron has no place to go since all levels that it could occupy are already filled, and the energy required to cross the forbidden gap, E_g, and reach the conduction band is much greater than the available thermal energy, that is, $E_g \gg k_B T$. On the other hand, in good electrical conductors such as metals, either the conduction band is filled to some extent or there is an overlap between it and the valence band (Figure 2.13b). There are then empty levels available for electrons to move into when the material is subjected to a voltage source. The material is classified as a semiconductor if the forbidden gap is small enough that electrons can be thermally excited from the valence to the conduction band, that is, $E_g \approx k_B T$ (Figure 2.13c). The material's conductivity then lies in between that of a good conductor and an insulator.

When an electron is raised from the valence band into the conduction band, a *hole* is created in the valence band (Figure 2.14). This has essentially the same characteristics as an electron, except that it has a positive charge. The hole also moves when subjected to an electric field. In fact, when a voltage is applied to the material, holes will flow toward the negative terminal, while electrons flow toward the positive terminal.

A pure semiconductor which contains no impurities is referred to as an *intrinsic semiconductor*. Such a semiconductor has an equal number of conduction electrons and holes, since the holes are

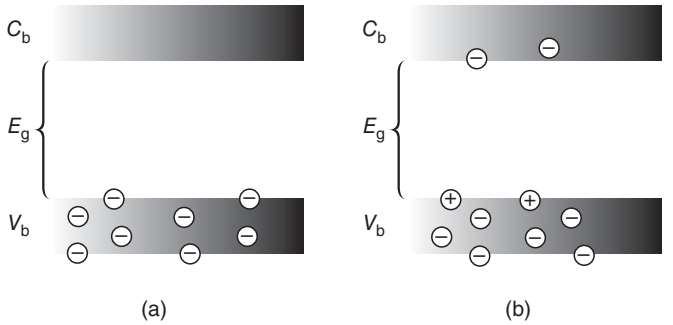


Figure 2.14 Creation of holes in a valence band. (a) Filled valence band. (b) Partially occupied conduction band with holes in valence band.

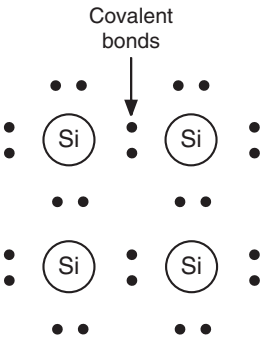


Figure 2.15 Covalent bonding in an intrinsic semiconductor.

created by the excitation of electrons to the conduction band. The number of holes in the valence band, and electrons in the conduction band (which are available for electrical conduction), increases with temperature, since the occupation of higher energy levels increases with temperature, according to Boltzmann's law. Thus, the conductivity of such materials increases with temperature (the opposite is true of metals). Examples of such class of materials are silicon and germanium, which are group IV elements of the periodic table, and each has four electrons in the outermost orbits of their atoms. Each atom shares its four outermost electrons with four other surrounding atoms to form covalent bonds (Figure 2.15), and since each bond involves two electrons, the outermost shell now has eight electrons altogether and is thus completely filled. It thus behaves as an insulator at absolute zero, with the conductivity increasing with temperature.

If now a group V element, say arsenic, which has five valence electrons, is added in small amounts as impurities to the germanium solid (a process called doping), four of its electrons will form stable bonds with the surrounding germanium atoms leaving the fifth atom not tightly bound and thus essentially free and mobile in the solid (Figure 2.16a). When a semiconductor is doped in this manner, it is said to be n-type, where the "n" indicates negative since there is an excess of electrons available for conduction. Arsenic is then a donor element. On the other hand, if the original solid is doped with a group III element, say gallium, which has three valence electrons, then there is a deficiency in bonding electrons, resulting in the creation of holes (Figure 2.16b). The resulting semiconductor is said to be p-type. The gallium atom is then an acceptor atom. It essentially takes an electron from the valence band, leaving holes behind.

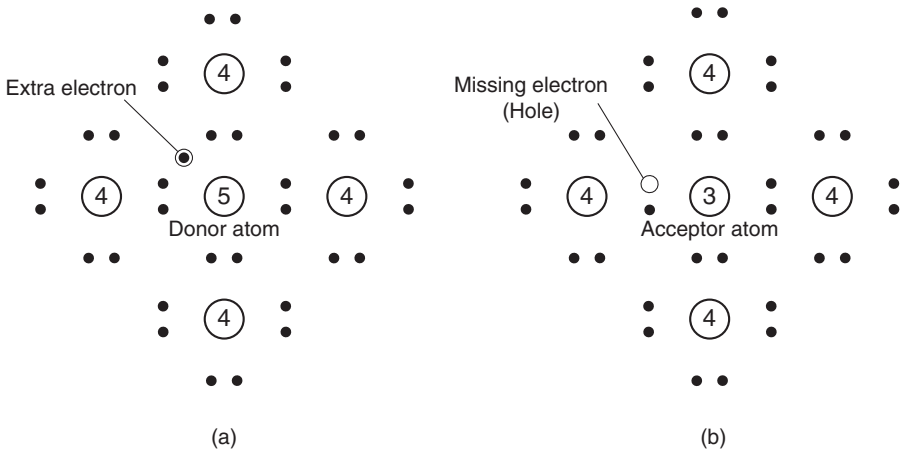


Figure 2.16 (a) n-Type semiconductor. (b) p-Type semiconductor.

2.3.2 Semiconductor Lasers

To understand the principle of operation of a semiconductor laser, first consider an intrinsic semiconductor at absolute zero temperature (Figure 2.17a), where the valence band (V_b) is completely filled (indicated by the shaded region). If electrons are now excited from the valence band into the conduction band (C_b), by say, pumping, these electrons rapidly drop down to the lowest states in the conduction band with a lifetime of about 0.1 ps. Likewise, the electrons in the valence band are also redistributed to fill the lowest levels that are empty, resulting in holes being created at the top of the valence band (Figure 2.17b). Under these circumstances, the semiconductor as a whole is not in thermal equilibrium, even though the redistribution process results in localized equilibrium in each band. Within each band, there is a quasi-Fermi level that separates the fully occupied and empty levels in that band. These are indicated by F_c and F_v for the conduction and valence bands, respectively, and their values depend on the number of electrons pumped to the conduction band. When the semiconductor as a whole is in equilibrium, $F_c = F_v$.

With the bottom of the conduction band being occupied by electrons while the top of the valence band contains holes, a population inversion then exists between the conduction and valence bands (Figure 2.17b). When the electrons in the conduction band drop down into the valence band and

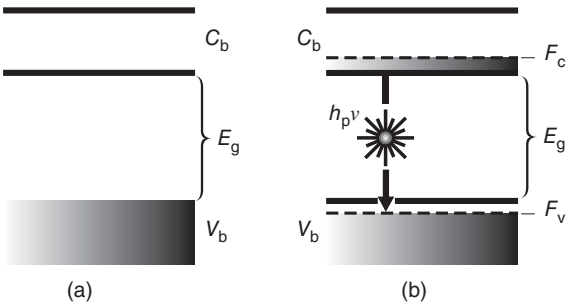


Figure 2.17 Basic concepts in semiconductor laser operation. (a) A completely filled valence band. (b) A population inversion resulting from a partially filled conduction band. Source: Svelto (2010)/Springer Nature.

recombine with the holes, either a photon (light wave) or a phonon (thermal wave) will be emitted depending on the material. In materials such as silicon and germanium where the radiation is emitted in the form of heat (phonons), no laser action can be obtained. Laser action can be achieved in materials like gallium arsenide (GaAs) where the radiation is in the form of light (photons).

The probability of occupation of any energy level, E, of each band is (for the conduction band, $p_c(E)$)

$$p_c(E) = \frac{1}{1 + e^{\left(\frac{E-F_c}{k_B T}\right)}} \tag{2.6}$$

(and for the valence band, $p_v(E)$)

$$p_v(E) = \frac{1}{1 + e^{\left(\frac{E-F_v}{k_B T}\right)}} \tag{2.7}$$

During pumping from energy level E_v in the valence band to energy level E_c in the conduction band, the rate of absorption depends on the following:

1. The density of incident radiation, $e(\nu)$.
2. The transition coefficient, B.
3. The probability that the valence band energy level is occupied, $p_v(E_v)$.
4. The probability that the conduction band energy level is empty, $1 - p_c(E_c)$.

It is therefore given by:

$$\frac{dN_1}{dt} = k_c B p_v(E_v)[1 - p_c(E_c)]e(\upsilon) \tag{2.8}$$

where k_c is a constant. Likewise, the rate of stimulated emission is given by:

$$\frac{dN_2}{dt} = k_c B p_c(E_c)[1 - p_v(E_v)]e(\upsilon) \tag{2.9}$$

Here, N_1 and N_2 are the number of photons absorbed and emitted, respectively.
Laser amplification will prevail only when $\frac{dN_2}{dt} > \frac{dN_1}{dt}$, that is,

$$p_c(E_c)[1 - p_v(E_v)] > p_v(E_v)[1 - p_c(E_c)] \tag{2.10}$$

Substituting for $p_c(E)$ and $p_v(E)$ from Eqs. (2.6) and (2.7) gives

$$F_c - F_v > E_c - E_v = h_P \nu \tag{2.11}$$

where ν is the emitted photon frequency, while E_c and E_v are the upper and lower laser levels, corresponding to energy levels in the conduction and valence bands, respectively.

When an electron undergoes a transition from an energy level E_c in the conduction band to an energy level E_v in the valence band, the energy associated with the transition is also greater than the energy gap, E_g. Thus,

$$F_c - F_v > h_p \nu = \frac{h_p c}{\lambda} > E_g \tag{2.12}$$

Laser action is more easily achieved in a semiconductor when using a p–n junction diode (Figure 2.18), which is obtained by joining an n-type to a p-type semiconductor. The energy levels

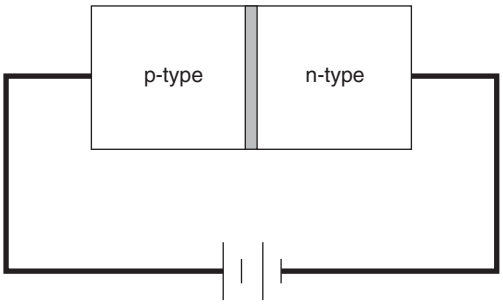


Figure 2.18 p–n junction diode (Forward bias)

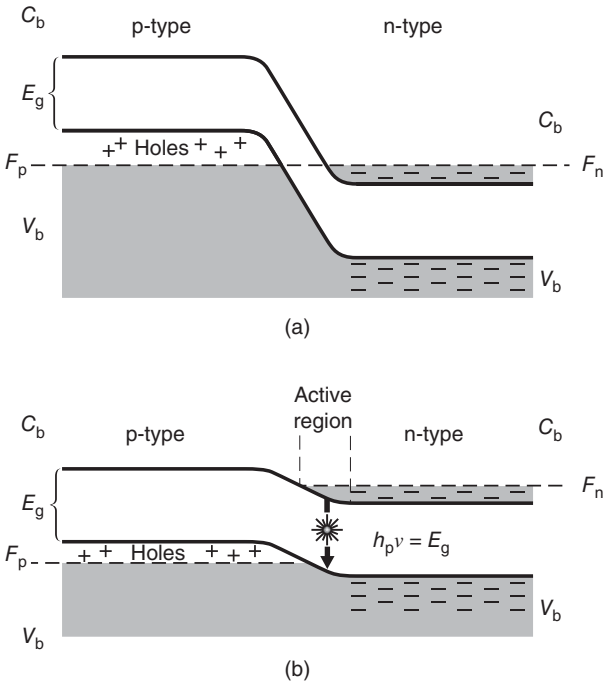


Figure 2.19 Energy levels of a heavily doped p–n junction diode. (a) Zero bias voltage and (b) forward bias. Source: From O'Shea et al. (1977)/with permission of Pearson.

for a diode are shown in Figure 2.19. The energy levels of the conduction and valence bands are different on the two sides of the junction, with those for the n-type being at a lower level than those for the p-type (Figure 2.19a). Population inversion is achieved if the diode is heavily doped with acceptors and donors. Under such circumstances, the Fermi level F_n for the n-type lies in the conduction band, while that for the p-type F_p lies in the valence band (Figure 2.19a). A portion of the conduction band in the n-type region is then occupied by electrons up to F_n, and in the p-type region, holes are added by the acceptors down to F_p.

When there is no voltage applied across the junction, excess electrons flow from the n-type to the p-type, setting up an electric field that opposes and thus stops the flow. The Fermi levels on the two sides of the junction, F_n and F_p, are then at the same level or have the same energy (Figure 2.19a).

If the junction is reverse biased by connecting the n-type portion of the junction to a positive electrode and the p-type portion to a negative electrode, the electrons flow toward the anode while the holes flow toward the cathode, until there are no more free electrons available. Current flow then stops. On the other hand, if the positive electrode is connected to the p-type side while the negative electrode is connected to the n-type side, forward biasing (Figures 2.18 and 2.19b), electrons flow into the junction as they move toward the anode. At the junction, they may recombine with holes, which are attracted in the opposite direction, to generate radiation. Current flow continues as the n-type region is supplied with more electrons, while more holes are created in the p-type side (as it is depleted of electrons) by the voltage source. This is the basic principle of a diode, conducting current in one direction, and preventing current flow in the other direction.

The flow of current due to forward biasing injects electrons from the conduction band of the n-type and holes from the valence band of the p-type materials into a narrow region in the immediate neighborhood of the p–n junction, called the depletion layer, which is only about 0.1 μm in thickness. That results in the creation of a population inversion in that region. Recombination of electrons and holes in this region generates radiation of energy equal to the bandgap energy. The recombination is equivalent to the de-excitation of an electron from the conduction band to the valence band. If the resulting emission is not stimulated, then the device is a light-emitting diode (LED). If, however, the applied current is high enough to exceed a certain threshold value, then a population inversion is achieved which results in stimulated emission and laser action. Above the threshold current, the beam intensity increases rapidly with increasing current and becomes more monochromatic, and its divergence is significantly reduced.

Normally, laser oscillation involves using two mirrors to form a cavity. In semiconductor lasers, this is achieved by cleavage of the material along natural crystalline planes to obtain highly reflecting and parallel surfaces at two ends of the crystal which are also perpendicular to the junction. The other two surfaces are left rough to suppress oscillation in undesired directions. Furthermore, since semiconductors have a very high index of refraction, about 3.6, the inherent reflectivity at the crystal–air interface is very high, about 35%, making it unnecessary to use reflecting coatings.

Semiconductor lasers have the advantage of being relatively small in size and easy to fabricate by mass production, thereby being low cost. Typical dimensions are $1\,\text{mm} \times 1\,\text{mm} \times 200\,\mu\text{m}$, with the thickness of the active region, where most of the oscillation is concentrated, being about 1 μm (Figure 2.20). However, due to diffraction, the beam itself is much wider, about 40 μm, extending

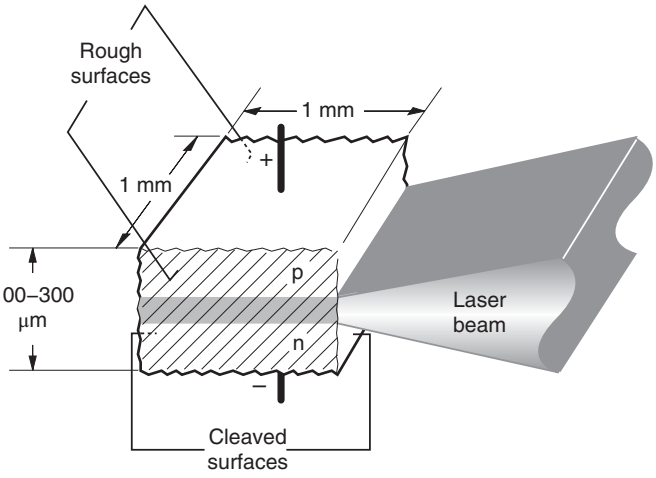


Figure 2.20 Schematic of a semiconductor laser.

away from the interface into the p and n regions. The extension of the beam into the regions away from the interface results in high cavity losses, since absorption is much greater than emission as one moves away from the interface. These losses result in high threshold current densities being required for laser oscillation.

The threshold current increases with temperature, and this is because $p_v(E_v)[1 - p_c(E_c)]$ increases, and $p_c(E_c)[1 - p_v(E_v)]$ decreases as temperature increases, and from Eq. (2.11), that means the oscillation gain decreases with increasing temperature. The relatively narrow width of the beam, about 40 μm, as it comes out of the optical cavity, results in high beam divergence, compared to other lasers, as it is diffracted at the cavity aperture. This may range from 5 to 15°. Also, since transition occurs between two broadbands, its monochromaticity is not as good as in other lasers. The direct conversion of electrical energy into optical energy, however, results in very high efficiencies for semiconductor lasers, up to about 50%. The output wavelengths range from 0.7 to 30 μm in the infrared region. Semiconductor lasers are particularly suited to optical fiber communications.

2.3.3 Semiconductor Laser Types

There are several forms of semiconductor lasers. These include the following:

1. Homojunction lasers.
2. Heterojunction lasers.
3. Quantum well lasers.
4. Separate confinement heterostructure lasers.
5. Distributed feedback lasers.
6. Vertical cavity surface-emitting lasers (VCSELs).

In the following sections, we shall provide highlights of the first three types that are listed earlier.

2.3.3.1 Homojunction Lasers Homojunction lasers are made out of a single semiconductor material such as GaAs. In such lasers, the high concentration of electrons and holes in the active region increases the refractive index of the material in that region. This results in the emitted light being internally reflected and thus confined to that region. However, the amount that is lost out of the region is significant enough to require high threshold current densities for oscillation. For example, the threshold current density at room temperature for GaAs is about 10^5 Amps/cm^2. Thus, homojunction lasers require cryogenic temperature environment for CW operation. At room temperatures, they can only be operated in pulsed mode. At liquid nitrogen temperature of about 77 K, the GaAs laser can be operated under CW conditions. The output wavelength at room temperature is about 0.9 μm, decreasing to 0.84 μm at 77 K, and can thus be tuned by varying the temperature.

2.3.3.2 Heterojunction Lasers Heterojunction lasers are made out of two or more materials. This significantly reduces the threshold current density to the extent that for some material combinations, CW operation can be achieved at room temperature. This is because the different materials used have different refractive indices and bandgaps. These differences cause radiation generated in the junction region to be reflected back into the region, resulting in higher efficiency. The difference in bandgap results in the carriers being confined to the junction region, and this reduces the threshold current density.

Heterojunction lasers can be one of two forms:

1. Single heterojunction.
2. Double heterojunction.

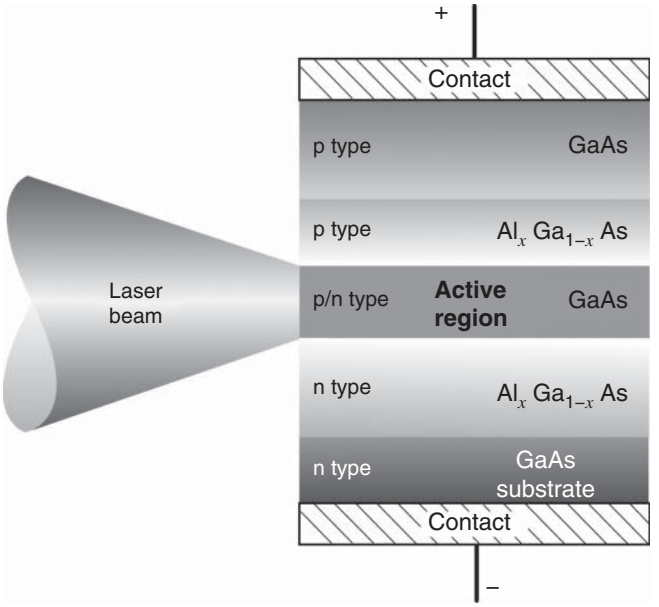


Figure 2.21 Schematic of a double heterojunction laser.

In a single heterojunction, the junction is formed between a p-type $Al_xGa_{1-x}As$ semiconductor and an n-type GaAs semiconductor, where x is a fraction. A typical double heterojunction has an active region of p-type GaAs which is a thin layer of about 0.1–0.5 μm, sandwiched between an n-type and a p-type $Al_xGa_{1-x}As$ semiconductor (Figure 2.21). There are thus two junctions between different materials. This reduces the threshold current density to about 10^3 Amps/cm^2, enabling room temperature CW operation to be achieved. This is due to a number of reasons:

(1) The difference in refractive indices of the materials (being about 3.4 for $Al_xGa_{1-x}As$ and about 3.6 for GaAs) confines the laser to the GaAs active region where there is high gain, thereby reducing losses due to absorption.
(2) The energy gap in the $Al_xGa_{1-x}As$ semiconductor is greater than that in the GaAs semiconductor. This acts as an energy barrier that reflects both holes and electrons back into the active region where they are confined, increasing their concentration and thus, gain in that region.

Since the output wavelength depends on the fraction x of Al in the semiconductor, the oscillation wavelength can be tuned by varying x. Peak pulse powers of 10 to 20 W can be obtained at room temperature with a pulse repetition rate of about 10^4 Hz, and a pulse duration of the order of a picosecond. Output powers of 5–100 mW can be achieved at room temperature under CW operation, but 5–10 mW is typical.

2.3.3.3 Quantum Well Lasers If the potential energy well of Figure 2.12a becomes so narrow that it confines the motion of electrons or holes to two dimensions (planar motion), then we have a quantum well. Confining motion to one dimension results in a quantum wire, and a quantum dot is obtained when motion is restricted in all three dimensions. The structure of the quantum well laser is similar to the heterojunction laser as illustrated in Figure 2.21, except that with the quantum well

laser, the middle layer is made thin enough to confine electron motion in this region to two dimensions. This improves the efficiency of the laser. Further improvements in efficiency can be obtained with further restrictions in motion to produce a quantum wire, and further yet, a quantum dot.

2.3.4 Low-Power Diode Lasers

The CW power output of diode lasers is comparable to that of He-Ne lasers. Thus, the two types of lasers are competitive in the 5–10 mW power range. Semiconductor lasers are much smaller, and their power requirements are lower (due to the higher efficiency), compared to He-Ne lasers. On the other hand, the output beam quality is much better for He-Ne lasers. Also, the 0.6328 μm output of He-Ne lasers is more sensitive to the human eye. However, they are more expensive than semiconductor lasers.

There have been significant improvements in the output powers of diode lasers in recent years. This is achieved by using stacks of diode lasers. In the next section, we present a brief background on high-power diode lasers.

2.3.5 High-Power Diode Lasers

The basic unit of high-power diode lasers is an individual diode laser with an output power of up to 100 mW (Figure 2.20). The power that is necessary for materials processing is acquired by combining the individual diode lasers in a linear setup to form an array. A number of arrays are further combined to form a laser bar of approximate dimensions 10 mm × 0.6 mm × 0.1 mm, and which is mounted on a heat sink or cooling device (Figure 2.22a). The output power of each bar is about 30 to 50 W. The individual bars are then stacked one on top of the other to form the laser diode stack (Figure 2.22b). The number of bars used determines the overall power output of the stack. Combining stacks enables output powers of up to 5 kW to be achieved.

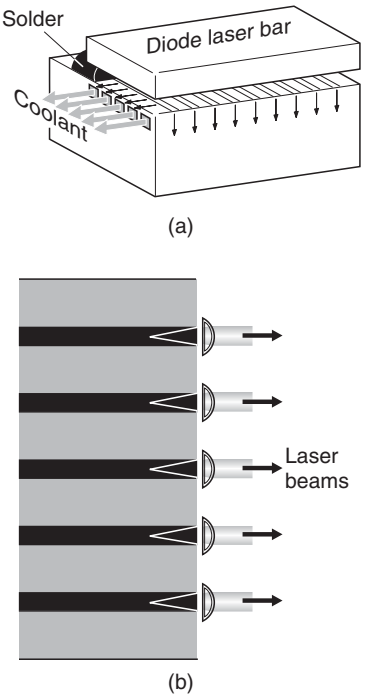


Figure 2.22 High-power diode laser construction. (a) Diode laser bar. (b) Stacked laser.

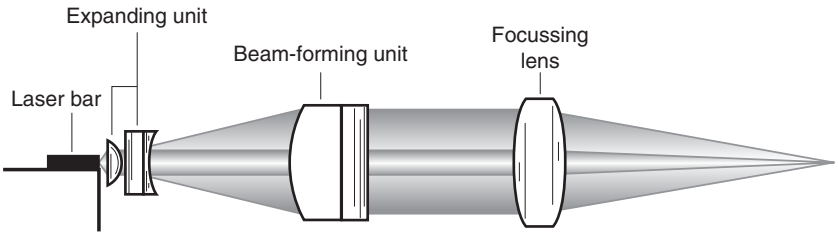


Figure 2.23 Setup of a stacked diode laser with beam transformation optics.

The lifetime of the laser diode stack is influenced by its operating temperature. More efficient operation is achieved at lower temperatures. Different methods have been developed for cooling the stack. One involves forcing a coolant through micro-channels etched into a silicon manifold layer, with each channel being approximately 25 μm in width and 150 μm in depth. Others involve the use of micro-fins or porous materials.

The beam from a diode laser is strongly divergent in the direction of the diode junction. Beam shaping therefore generally starts with collimating the beam in that direction. As an example, the divergence angle of an InGaAs-based diode laser along the fast axis is about 45–60°, while that along the slow axis is about 12°. Microlenses may be mounted directly in front of the diodes to collimate the fast axis and reduce the divergence angle in that direction to about 0.6°. Since the emitting aperture of the laser is rather large, about $10 \times 15\ \text{mm}^2$, and the divergence differs in different directions, cylindrical lenses are necessary to collimate the slow-axis divergence, before using an ordinary lens to focus it down to about $8.56 \times 0.62\ \text{mm}^2$ (Figure 2.23), which would be desirable for heat treatment.

The advantages of high-power diode lasers compared to the conventional CO_2 and Nd:YAG lasers include the following:

- Compact size.
- High efficiency (over 30%).
- Almost maintenance-free operation.
- Fiber optic beam guiding capability.
- Higher absorption at metallic surfaces.

The disadvantages include the following:

- Low power.
- Poor beam quality.

2.3.6 Applications of High-Power Diode Lasers

Early applications of diode lasers were limited to pumping of solid-state lasers, where the diode laser substitutes for the lamp. This increases efficiency, beam quality, lifetime, and reliability of the solid-state laser. Other applications in materials processing were in areas such as surface treatment, soldering small components, and welding plastics. With increased output powers of up to about 5 kW, diode lasers have also found application in laser welding but are mainly limited to conduction mode welding of thin sheets due to the relatively low-power densities of up to $3–4 \times 10^5\ \text{W/cm}^2$.

A class of high-power diode lasers (with output powers in the hundreds of watts) are effective for welding high reflectivity metals such as copper. These have been termed “blue lasers” for the visible

wavelength radiation that they generate. In using infrared lasers to weld the high reflectivity metals, a high initial power intensity is necessary to initiate the keyhole, since the absorption at the infrared wavelengths is low (Figure 8.5). Once the keyhole is initiated and absorption dramatically increases, it becomes necessary to control the power and/or speed to prevent spatter and/or porosity formation. However, absorption of visible wavelength radiation by copper is much higher than for infrared radiation (Figure 8.5). Thus, with the blue laser, the high initial power is not required, reducing the risk of spatter and porosity formation. The process window is also broadened.

One example blue laser might deliver 150 W of 0.450 μm beam through a 200 μm diameter optical fiber. Another might be a 500 W beam that generates a 1.4 MW/cm^2 power density, resulting in 500 μm penetration in copper at a speed of about 20 mm/s.

Since the output of the diode laser is usually either rectangular or square in shape, special optics are required to transform the beam from the linear emission characteristics of the laser bars to a round (or square) shape for coupling into an optical fiber.

2.4 NEW DEVELOPMENTS IN INDUSTRIAL LASER TECHNOLOGY

There have been significant advances in laser technology in recent years. In this section, we discuss the principal developments in laser technology that are useful for materials processing. We focus, in no particular order, on:

1. Slab lasers.
2. Disk lasers.
3. Ultrafast lasers.
4. Fiber lasers.

2.4.1 Slab Lasers

The active medium for solid-state lasers is traditionally shaped in the form of a cylindrical rod, with pumping and cooling occurring in the radial direction, while the beam propagates axially. However, with this configuration, a radial temperature gradient is induced in the rod. This leads to thermal focussing effects that cause the beam to be distorted at high laser power levels. This makes it difficult to obtain a low-order mode of operation or high beam quality at such power levels. This phenomenon is referred to as *thermal lensing*.

The use of a slab geometry for the active medium enables the beam quality to be significantly improved. The slab configuration is shown schematically in Figure 2.24. Both pumping and cooling

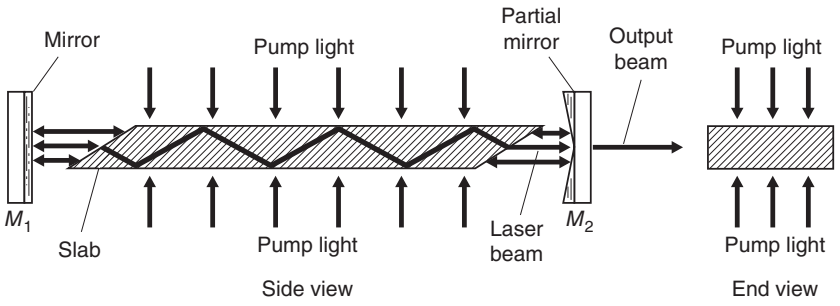


Figure 2.24 General schematic of a slab laser.

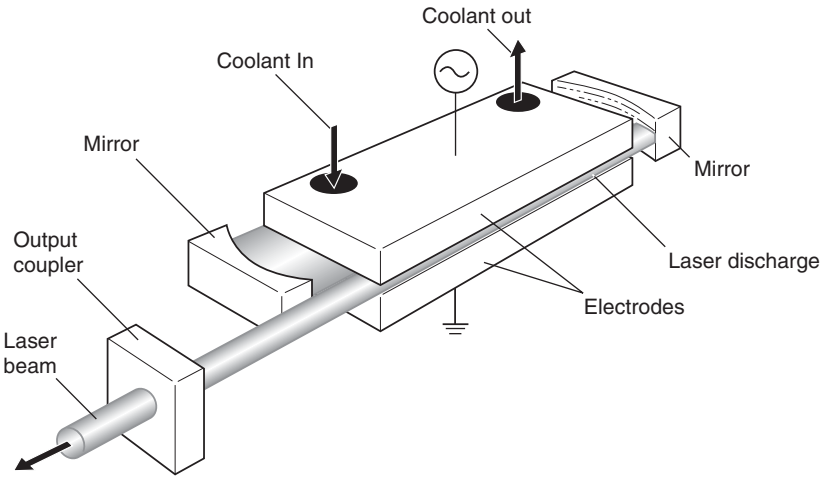


Figure 2.25 Schematic of a CO_2 slab laser. Source: With permission of Coherent, Inc.

occur in a direction normal to the flat faces of the slab and is thus one-dimensional, resulting in only one-dimensional temperature gradient. The beam is subjected to total internal reflection, thus encountering the entire array of temperature variation in the medium. This minimizes thermal lensing effects and improves beam quality.

Slab lasers can have either solid-state or gas active media, and Figure 2.25 illustrates a CO_2 slab laser. Here, the active medium is excited using two parallel rf electrodes, which are water-cooled. The system is thus said to be diffusion-cooled. As a result, the conventional gas circulation systems involving roots blowers or turbines are not required.

2.4.2 Disk Lasers

Disk lasers are also solid-state lasers and were developed to circumvent essentially the same problems for which slab lasers were developed – thermal lensing at high-power levels. The principle of the disk laser is illustrated in Figure 2.26. The active medium, which is shaped in the form of a thin

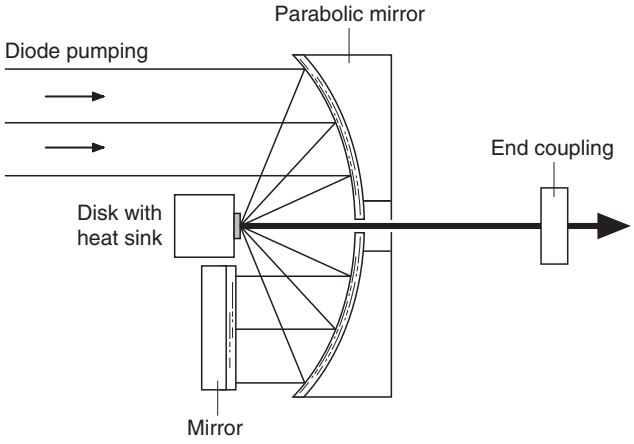


Figure 2.26 Schematic of a disk laser.

disk, is pumped using a diode laser. Parabolic mirrors are used to direct the diode laser output to one face of the disk. Excess heat that is generated in the disk is removed by a heat sink that is attached to the other face of the disk, resulting in efficient cooling. Since the resulting thermal gradients are parallel to the laser beam axis, the thermal lensing effect is minimized. Beam quality is thus significantly improved and is about four times better than that of a comparable diode-pumped rod laser. For example, a beam quality of $M^2 < 1.5$ can be obtained at a laser power of about 1 kW. Output powers of 6 kW per disk have been achieved, and multi-kilowatt output in excess of 20 kW can be obtained by connecting disks in series.

The higher beam quality enables a smaller focused beam size to be achieved. This results in a much higher power density, which enables higher processing speeds to be achieved. For example, cutting speeds of about 60 m/min have been achieved in cutting 0.2 mm-thick aluminum sheets. Welding speeds of over 50 m/min have also been achieved in 0.2 mm-thick stainless steel sheets.

Advantages of the thin disk laser system may thus be summarized as:

1. Reduced thermal lensing effect.
2. More efficient cooling of the laser.
3. High beam quality, and thus high-power density.
4. Higher laser efficiency.

2.4.3 Ultrafast (Femtosecond) Lasers

The laser systems discussed thus far are capable of generating either CW or pulsed beams with pulse duration of up to hundreds of picosecond duration by either Q-switching or mode locking. Shorter laser pulses of pulse duration down to femtoseconds are obtained using ultrafast (or ultrashort pulse) lasers. These enable much higher peak intensities to be obtained. For example, a peak intensity of 10^{15} W/cm^2 can be obtained from a 100 fs (10^{-13} s) laser pulse with a pulse energy of 0.33 mJ when focused to a 20 μm diameter. To achieve the same peak intensity with a 10 ns laser pulse, the energy content would have to be 100 J.

Different methods have been developed for generating ultrashort pulse beams. In this section, we discuss one of the basic principles, chirped-pulse amplification (CPA). The spectrum of a laser beam is inversely proportional to its pulse duration. Thus, a laser gain medium with a broad, continuous emission spectrum is necessary for generating an ultrashort pulse. The emission spectrum of traditional media such as CO_2 and Nd:YAG is such that they cannot support pulse duration down to the femtosecond range. Nd:YAG lasers, for example, can only generate pulse duration down to 30 ps. Typical materials that are used as active media for ultrafast lasers include Ti:sapphire (6 fs), Nd:glass (100 fs), Yb:glass, Yb:YAG, Cr:YAG, and dyes.

The generation of ultrashort pulses starts with mode locking. This generally results in a pulse repetition rate of about 100 MHz, with a pulse energy of the order of nanojoules. For a number of applications, it is necessary to amplify the pulse to the microjoule and millijoule range. However, due to the high intensity of the pulses, they cannot be directly amplified, since this would damage the optical components. The technique of CPA is one approach that is used to overcome this limitation. The CPA principle is illustrated in Figure 2.27. The peak power of the pulse generated by the oscillator is first reduced by stretching the pulse duration by a factor of 10^3 to 10^4 to make it longer. The stretched pulse is then amplified and then finally recompressed back to its short pulse state using a compressor. To avoid damage to the compressor, the laser beam diameter is expanded before being sent to the compressor.

Pulse stretching is normally achieved using a pair of optical gratings which are not parallel, or by using optical fibers. For pulses of relatively narrow bandwidths, a fiber optic cable is used. This first

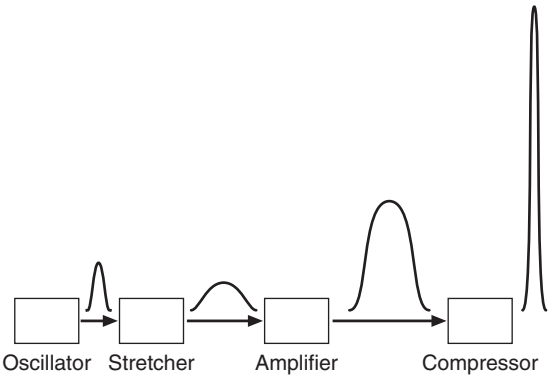


Figure 2.27 Principle of chirped-pulse amplification for ultrashort pulse generation. Source: Adapted from Liu et al. (1997).

broadens the pulse bandwidth by about 50 times and then stretches the widened pulse, usually about 3 to 5 times. Optical gratings are used for pulses that have large bandwidths. The principle of the latter stretching process is illustrated in Figure 2.28a. The width of a pulse that has a broadband spectrum is changed as a result of the dispersive properties of the gratings. In the original pulse, all the frequency components travel together. However, the gratings disperse the different frequency components into different directions, with each one having a different optical path length. And since the frequency component with the shorter optical path length emerges earlier than the one with the longer optical path length, the pulse becomes stretched in time.

For compression, the gratings are arranged to be parallel (Figure 2.28b). The separation between the gratings determines the amount of compression that is achieved.

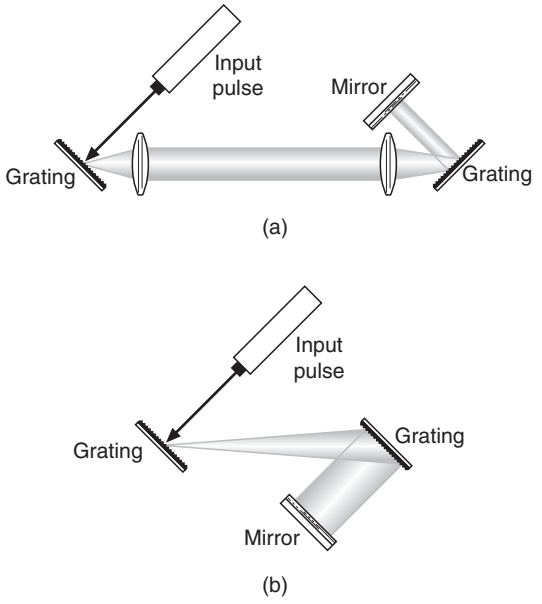


Figure 2.28 Chirped-pulse amplification. (a) Pulse stretching using an antiparallel grating pair. (b) Pulse compression using a parallel grating pair. Source: Adapted from Liu et al. (1997).

A typical ultrafast laser may have a Ti:sapphire active medium operating in a wavelength range of extreme ultraviolet to mid-infrared with a pulse duration of less than 10 fs and generates peak pulses of the order of terawatts at a pulse repetition rate of 1 MHz, resulting in an average power over 100 W.

Some applications of ultrafast lasers are discussed in Sections 9.3.1 (Micromachining) and 14.2 (Nanotechnology Applications).

2.4.4 Fiber Lasers

As the name implies, a fiber laser uses an optical fiber (Section 3.7.2) with a core that is doped with a rare-earth element as the active medium. It may therefore be considered as a form of solid-state laser. Fiber lasers were first produced in the mid 1960s, with output powers of the order of milliwatts at a wavelength of about 1 μm at that time. Current versions have output wavelengths ranging from ultraviolet through mid-infrared, with CW output powers up to 100 kW.

The basic structure consists of a single-mode or multimode core which contains the doping element and is surrounded by an undoped outer core (also known as inner cladding) (Figure 2.29). This is further surrounded by an outer cladding of even lower refractive index (double-cladding).

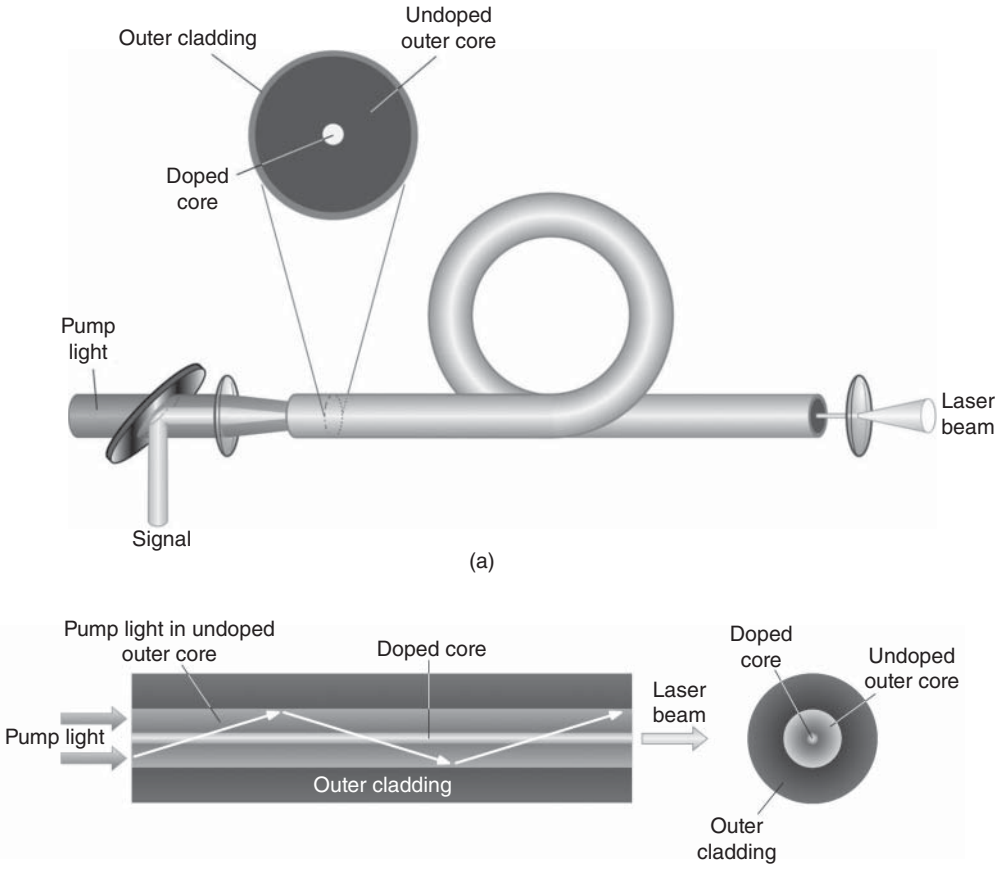


Figure 2.29 Basic structure of a fiber laser. (a) Overall setup. (b) Section view of the fiber. Source: From Hecht (2012)/Endeavor Business Media.

The core diameter is of the order of 40 μm, while the inner cladding is about 1 mm in diameter. The fiber can be as long as 10 m.

The challenge in achieving high beam quality (single mode) with high output power is one of increasing the core radius without having the higher-order modes propagating. However, as can be deduced from Eq. (3.22), for a given normalized frequency which results in single-mode operation, a larger core radius can be obtained only by decreasing the numerical aperture (NA). However, this is constrained by current fabrication technology, which limits the NA value to approximately 0.06, corresponding to a maximum core diameter of approximately 15 μm in the 1 μm wavelength range. One common approach is to induce bending losses, which are more significant for the higher-order modes than the fundamental mode, by proper coiling of the fiber.

In the simplest case, the doped optical fiber core, with the ends cleaved and butted with dielectric mirrors (multiple thin layers of different transparent optical materials), constitutes the laser cavity. In practice, however, fiber Bragg gratings are used for the end mirrors, since these are easier to mass produce. The gratings are designed to reflect light in a specific wavelength range and allow all other wavelengths through.

In a fiber Bragg grating, stripes are introduced into a section of the glass such that the effective refractive index in that region is varied (periodically or aperiodically). Whenever the light crosses the boundary between two sections of the material with different refractive indices, a small portion is reflected back. Thus with a sufficient number of stripes, the grating can be an efficient mirror.

Common elements used for doping include erbium (Er), holmium (Ho), neodymium (Nd), thulium (Tm), and ytterbium (Yb). Neodymium and ytterbium are used extensively for high-power applications, with Yb resulting in the most efficient and powerful laser. The host medium is usually silica. The laser output wavelengths corresponding to the various active media are shown in Table 2.1, together with the normal pump wavelengths. The difference between the pump and output wavelengths is referred to as the quantum defect, and that determines the excess heat produced by the laser. Since the difference between the two is small for fiber lasers, the quantum defect heating and resulting thermal load are relatively low. That is especially the case with Yb lasers and is one of the reasons for the high beam quality and high efficiency of fiber lasers.

2.4.4.1 Pumping Pumping of the fiber laser is normally done optically, using a diode laser, either longitudinally (end pumping) or transversely (side pumping). In either case, the source may be a single diode, an array, or several individual diodes. Since the diode laser output is low brightness (low quality), launching it directly into the fiber core is difficult. It is therefore launched into the inner cladding (pump core) (Figure 2.30), which has a much larger cross-sectional area (about 100–1000 times that of the core). Thus, higher power diode lasers can be used for pumping. The pumping beam is internally reflected repeatedly, and part of it is absorbed each time it crosses the fiber core. This absorption mechanism, and the long length of the fiber, enables high gains to be achieved, since the laser medium gain is given by the product of the pump light intensity and the distance over which the pump light interacts with the active medium. The result is a high-brightness, high-power output radiation. With side pumping, the diode output is first fed into a coupler, and at an angle to the fiber axis (Figure 2.31).

Table 2.1 Output wavelengths for common fiber lasers.

Active medium	Pump band (μm)	Output wavelength (μm)
Neodymium	0.808	1.064–1.088
Ytterbium	0.900–0.980	0.970–1.200
Erbium	0.980, 1.480	1.54–1.62
Thulium	0.793	1.750–2.100

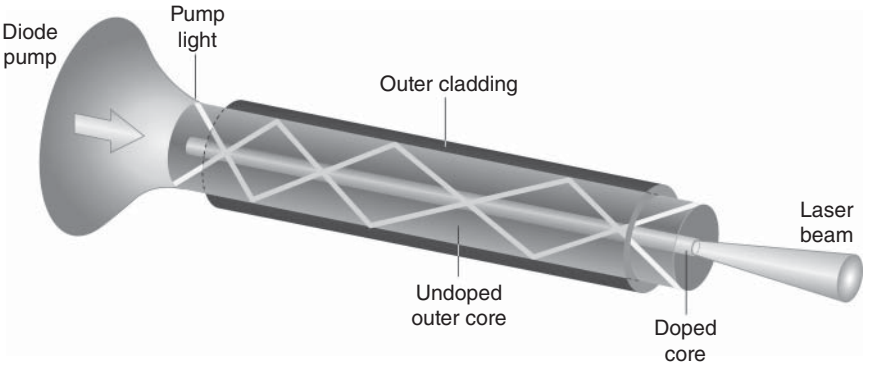


Figure 2.30 Pumping a fiber laser. Source: Richardson et al. (2010)/Optical Society of America.

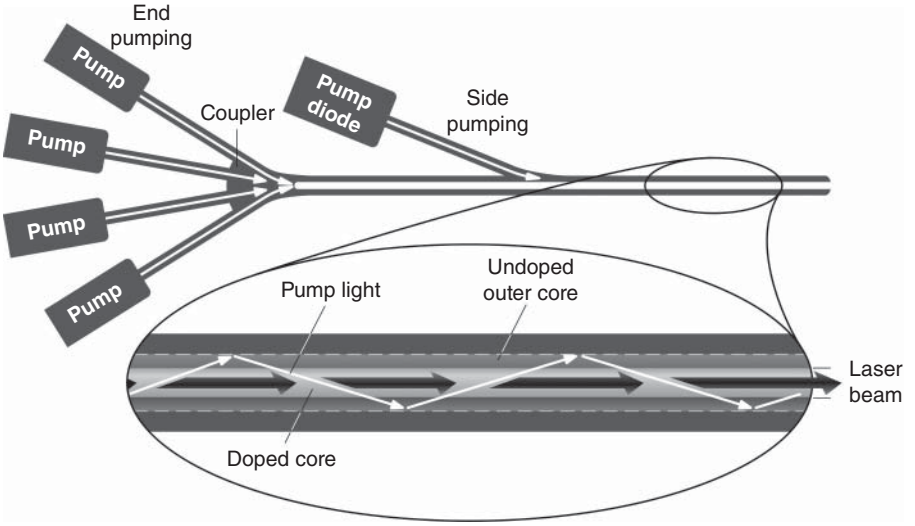


Figure 2.31 Illustration of end and side pumping techniques. Source: From Hecht (2012)/Endeavor Business Media.

With cylindrical double-clad fibers, the reflected light rays do not always pass through the core, and that reduces the pumping efficiency. To ensure that all modes of the inner cladding are coupled to the core, the optical fiber is either coiled or the cladding is made noncircular in cross section. Even though nonsymmetric shapes such as the D-shape were initially investigated, symmetric shapes such as octagonal or hexagonal are preferable since they are much easier to splice to another fiber, and also to fabricate.

Transitions in fiber lasers are characteristic of quasi-three-level systems. Pumping is from a ground state to an upper level. Laser transition occurs from the lowest part of the upper level to the ground states. The output may be CW or pulsed.

2.4.4.2 Pulsed Output Pulsed outputs can be obtained by

1. Modulation of the pump laser to obtain long-pulse (millisecond duration) output.
2. Q-switching to generate pulses ranging from nanosecond through microsecond duration.

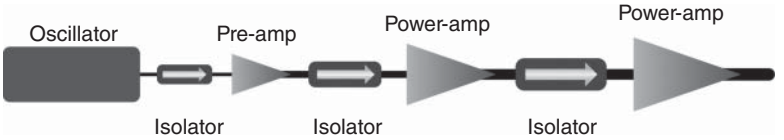


Figure 2.32 Schematic of a three-stage amplifier MOPA.

3. Mode-locking to generate pulses of the order of 100 femtosecond duration.
4. CPA (Section 2.4.3) for even shorter pulses.

Using the CPA technique, pulse peak powers in excess of 1 GW have been achieved with an average power of about 800 W, and pulse duration can be as short as 10 femtoseconds.

High power/energy pulses in the nanosecond range have been obtained by using high-gain fiber master oscillator power amplifiers (MOPAs) and essentially consists of a master oscillator and a number of amplifiers connected in series (Figure 2.32). The master oscillator generates a low-power output which is amplified sequentially to the desired power level. The drawback of this approach is that nonlinear effects that result from the long length and small diameter of the fiber can limit the output power produced. However, this is sometimes circumvented with the use of large mode area (LMA) active fibers, enabling higher outputs to be obtained. The greater absorption per unit length afforded by LMA fibers enables shorter fibers to be used.

Pulsed outputs have been obtained at wavelengths ranging from 1 to 3 μm using a variety of techniques. These are extensively used in materials processing but have found application in other areas such as nonlinear optics, environmental monitoring and pollution control, food quality control, and noninvasive disease diagnosis. In materials processing, high-power (energy) pulsed nanosecond lasers are used for laser marking and engraving in a variety of materials such as metals, plastics, ceramics, organics, and paper. High-power ultrafast fiber lasers are used for micromachining, welding, cutting, drilling, surface texturing, and microcladding.

2.4.4.3 CW Output Single-mode fibers enable diffraction-limited output beams to be generated but result in relatively lower output powers, while multimode fibers are able to produce much higher output beams of relatively lower quality. The CW output power of single-mode fiber lasers is typically up to hundreds of watts and, when collimated to 5 mm diameter, can result in a full angle divergence of about 0.13 mrad. The highest power achieved thus far for a single-mode laser with a LMA is 10 kW. This was achieved by tandem pumping, where one or more fiber lasers are used to pump another fiber laser. In this case, Yb-doped fiber lasers, each with an output of several hundred watts, were used to pump another at 1.018 μm pump wavelength, generating an output beam at 1.070 μm. The resulting quantum defect is smaller than 5%, which is about half what would have been obtained with diode pumping.

Higher power levels can also be achieved by combining the output from several fiber lasers. This process can be coherent or incoherent. For example, the outputs of 10 fiber lasers can be integrated to produce a 1 kW output system. A 33 kW output laser has been produced by incoherently combining the outputs of six 5.5 kW fiber lasers. Powers higher than 100 kW CW have also been achieved with multimode fibers by double-cladding.

2.4.4.4 Power Scaling Scaling of power in fiber lasers is limited by nonlinear effects, which depend on (proportional to) the beam intensity in the fiber core and the fiber length. One source of nonlinear effects is third-order susceptibility. This makes the refractive index of the medium intensity-dependent, thereby introducing a self-induced phase shift (self-phase modulation).

Second-order effects cancel out in silica fibers, since silica is a symmetric molecule. Fortunately, self-phase modulation effects are so low that they can be neglected in CW fiber lasers. Another source of nonlinearity is stimulated inelastic scattering. Above a threshold radiation level, stimulated Raman scattering (SRS) or stimulated Brillouin scattering (SBS) results in part of the radiation energy being transferred to the glass host. SRS involves excitation of optical phonons, while SBS involves excitation of acoustical phonons, and its effect is much smaller than that of SRS.

The threshold power for SRS is given by:

$$q_{\mathrm{RS}} = \frac{16A_{\mathrm{e}}}{g_{\mathrm{RS}}L_{\mathrm{e}}} \tag{2.13}$$

where A_{e} is the effective fiber mode area, L_{e} is effective length of the fiber, and g_{RS} is the peak Raman gain coefficient. It is 1×10^{-13} m/W in fused silica at a pump wavelength of 1 μm.

A similar relationship exists for Brillouin scattering. Eq. (2.13) indicates that the threshold power at which nonlinear effects become significant decreases as the fiber length or beam intensity increases, or as the mode-field area decreases. SBS is significant enough to restrict the performance of single-frequency fiber lasers, while for spectral bandwidths greater than 0.5 GHz, power scaling is constrained by SRS.

2.4.4.5 Applications of Fiber Lasers The higher-quality beam associated with single-mode fibers makes them suitable for applications such as laser cutting or transmission through the atmosphere. On the other hand, the higher power output of multimode fibers makes them suitable for laser welding and heat treatment, which normally require a wider beam area.

High-power Yb fiber lasers operating in the 1 μm wavelength range with CW output have become the workhorse of the laser industry for processing metals, since this wavelength range is efficiently absorbed by metals. Example applications include using a 2 kW fiber laser for lap-welding a zinc-coated steel sheet of thickness 1.2 mm at 5 m/min or cutting a 4 mm-thick coated steel at 10 m/min, producing clean sharp edges. In another example, a 6 mm-thick 6056 aluminum alloy was welded with a 6.9 kW laser at 3 m/min. Other applications include laser drilling, cladding, and cleaning.

Significant improvements in output power are being obtained for fiber lasers operating in the 1.5, 2, and 3 μm wavelength ranges. These output wavelengths have the advantage that they are eye-safe, since they are easily absorbed by water in the vitreous humor, and thus unable to reach the retina where damage can be done. Thulium (Tm) and Holmium (Ho) lasers operate in the 2 μm range and are effective for welding, drilling, and cutting nonmetals such as plastics which are transparent around the 1 μm range but absorb radiation around 2 μm efficiently. These lasers have been effectively used in welding plastics to metals and have also found significant use in medical applications such as ablating urinary or canine prostate tissue, since the water in the tissues is very effective in absorbing the 2 μm wavelength radiation. For polymers, an ultraviolet fiber laser generating 10 μJ laser pulses at 200 kHz (2.0 W) can produce a light black bar code at a marking speed of 1 m/s on ABS plastic.

2.4.4.6 Advantages of Fiber Lasers Fiber laser advantages include the following:

1. Compact size, since the fibers can be packaged in a coiled configuration. For example, a 4 kW fiber laser's footprint is about 0.5 m^2 as compared to 11 m^2 for an Nd:YAG laser that is lamp-pumped. An 8 kW fiber laser's dimensions may be about $1.5 \times 0.8 \times 1.5$ m.
2. High brightness output beam, which is diffraction-limited and therefore of high quality, with M^2 value of 1 to 2 for the single-mode low-power lasers. The M^2 value for high-power multimode operation typically ranges between 7 and 10. The high beam quality enables a 1 kW

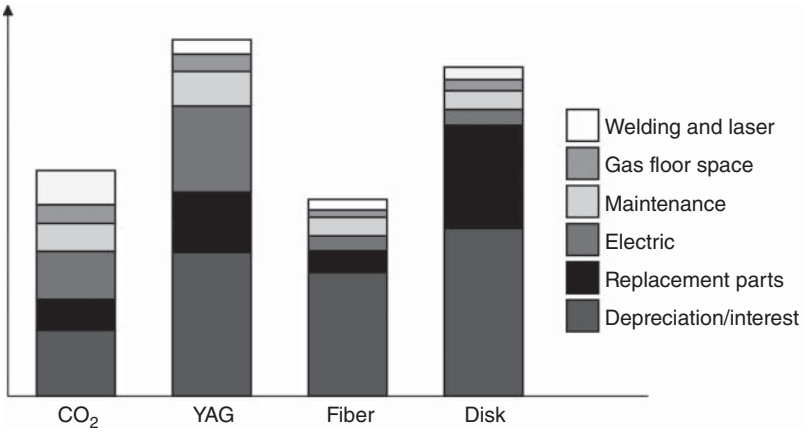


Figure 2.33 Relative operating costs for industrial lasers. Source: Adapted from Quintino et al. (2007).

laser to be focused to a 25 μm beam radius using a lens of focal length 100 mm, while an 8 kW beam is focused to a 300 μm beam radius using a lens of focal length 250 mm.

3. Ability to generate ultrashort pulse laser beams and/or wide tunable range, due to the broad emission spectrum of the glass gain medium.
4. Relatively high output efficiency of about 25–50% compared to about 10–30% for CO_2 and 1–3% for Nd:YAG lasers. Thus, less of the input energy is wasted. Furthermore, since the fiber has a large surface area-to-volume ratio, any excess heat is efficiently dissipated over the long length of the fiber. There is therefore little risk of the fiber overheating and affecting the beam quality. The optical to optical conversion efficiency typically ranges from 60% to 70% and can be as high as 95% for erbium. This coupled with the high efficiency of diode lasers (about 50% electrical to optical) is what results in the relatively high efficiency of fiber lasers.
5. Near-infrared wavelength of the laser enables relatively high absorptivity to be achieved in metals.
6. Relatively lower operating costs. Costs associated with using some of the major laser types are shown in Figure 2.33. These costs are averaged over an 8-year period and show the operating costs for fiber lasers to be lower than for CO_2, Nd:YAG, or disk lasers.

2.4.4.7 Disadvantages of Fiber Lasers Fiber laser disadvantages include the following:

1. Difficulty in aligning the pump laser output to single-mode core optical fibers.
2. Potential for fiber damage at high powers.
3. Need to use long cavity lengths, since the pump absorption per unit length is limited.

2.5 SUMMARY

In this chapter, we have discussed some of the major types of lasers, focusing on those that are either currently used, or have the potential to be used for manufacturing applications. The discussion started with solid-state lasers, with a focus on the Nd:YAG laser (which is four-level and is extensively used in manufacturing applications). The Nd:YAG laser is capable of transmission using a fiber optic

cable and can be used in either the pulsed or CW mode, with CW power levels up to 6 kW. The Nd:Glass laser is able to deliver higher output peak power pulses than the Nd:YAG laser. However, it has a more limited use in manufacturing because of its inability to deliver CW output power for any reasonable period of time.

The discussion on solid-state lasers was followed by gas lasers, where the focus was on neutral atom (He-Ne), ion (argon ion), and molecular (CO_2 and excimer) lasers. The low-power CW output of the He-Ne laser makes it suitable for position sensing and alignment, among other applications. The CW power output of the argon ion laser is higher than that of the He-Ne laser, but much lower than can be obtained from the CO_2 laser. The CO_2 laser, which is the other laser that is extensively used in processing, is much more efficient than the Nd:YAG laser and generates much higher CW output powers, as well as higher beam quality. However, the longer wavelength output does not lend itself to fiber optic cable transmission. The CO_2 lasers that are most extensively used in industry are the fast axial flow and transverse flow lasers, which produce the highest power outputs. The excimer laser, which normally generates pulsed beams in the ultraviolet wavelength range, is beginning to see increasing use in manufacturing.

Diode lasers constitute one of the fastest growing laser type in materials processing. Even though the power output of an individual diode laser is relatively small, stacking of the individual lasers enables high output powers suitable for materials processing to be achieved. The near-infrared wavelength of the diode laser makes it suitable for transmission using fiber optic cables. The major drawback of this type of laser is the high beam divergence, which requires special optics for collimating and focusing the beam. Even then, it is limited to processes that require a relatively wide beam area, such as surface modification, and conduction mode welding.

Following diode lasers, slab lasers were discussed. These can have either solid state or gas as the active medium. However, the pumping configuration enables higher-quality output beams to be obtained compared to the beams generated using traditional pumping techniques. This was followed by a discussion on disk lasers, which produce even better beam quality because of reduced thermal lensing effects. Ultrafast lasers were then discussed. Even though these are also solid-state lasers, they were considered separately because of the special characteristics imparted by the ultrashort pulse duration in materials processing, enabling materials to be ablated or removed with minimal thermal damage to the surrounding material.

Finally, fiber lasers, where the active medium is a doped optical fiber, were discussed. These have the advantage of compactness, coupled with high beam quality, and very high output powers.

APPENDIX 2.A

List of symbols used in the chapter

Symbol	Parameter	Units
C_b	Conduction band	—
e^-	Electron	—
E_c	Energy level in conduction band	J
E_g	Energy of forbidden gap	J
E_v	Energy level in valence band	J
F_c	Energy of Fermi level in conduction band	J

Symbol	Parameter	Units
F_n	Fermi level for n-type semiconductor	J
F_p	Fermi level for p-type semiconductor	J
F_v	Energy of Fermi level in valence band	J
He^*	Excited helium atom	—
k_c	Constant	$/m^3\text{-}Hz^2$
k_m	Constant	—
n_m	Number of periods in magnetic assembly	—
p_c	Probability of occupation of conduction band	—
p_v	Probability of occupation of valence band	—
V_b	Valence band	—
γ_e	Ratio of electron energy to its energy at rest	—
λ_m	Oscillatory period of magnet	m
$\Delta\nu_o$	Linewidth of radiation	Hz

APPENDIX 2.B

Comparison of common industrial lasers

Characteristic	CO_2	Nd:YAG	Excimer	Diode	Fiber
Wavelength (μm)	10.6	1.06	0.125–0.351	0.6–10	0.5–3
Divergence (degrees) 0.1–0.2	0.06–1.15	0.06–0.3	5–30	<7	
FWHM (GHz)	0.1–4	195	8–4000	1875	1600
Output power CW (W)	30×10^3	8×10^3	5–200	$(1\text{–}100) \times 10^{-3}$	$(1\text{–}500) \times 10^3$
Peak power (MW)	0.030	50	10–100	10^{-5}–10^{-6}	1
Overall efficiency (%)	10–30	1–3	1–4	30–50	25–50
Focused beam intensity (W/cm^2)	$10^{6–8}$	$10^{5–9}$	—	$10^{3–5}$	$10^{7–9}$
Laser head size (10^{-3} m^3)	10^3	10^2	—	1	6
Maintenance periods (Hours)	2000	200 (Lamps)	—	Minimum	Minimum
Pulsed or CW	CW/pulsed	pulsed/CW	pulsed/CW	CW/pulsed	CW/pulsed
Pulse rate (Hz)	1–5000	1–100	1000	12×10^6	50–>10^6
Pulse duration (s)	10^{-4}	10^{-3}–10^{-8}	10^{-9}	10^{-12}	10^{-9}–10^{-13}

APPENDIX 2.C

Wavelength range of common lasers (Table 1.1, Henderson, 1997)

Type	Wavelength (μm)
Krypton-fluoride excimer	0.249
Organic dye	0.3–1.0 (tunable)
Xenon-chloride excimer	0.308
Helium-cadmium	0.325, 0.442
Krypton ion	0.335–0.800
Nitrogen	0.337
Argon ion	0.450–0.530 (0.488 and 0.515 strongest)
Copper vapor	0.510, 0.578
Helium-neon	0.543, 0.633, 1.15
Gold vapor	0.628
Titanium sapphire	0.660–1.06 (tunable)
Semiconductor (GaInP family)	0.670–0.680
Ruby	0.694
Alexandrite	0.700–0.830 (tunable)
Semiconductor (GaAlAs family)	0.750–0.900
Nd:YAG	1.064
Semiconductor (InGaAsP family)	1.3–1.6
Carbon monoxide	5.0–6.0
Carbon dioxide	9.0–11.0 (main line 10.6)
Fiber lasers	UV to mid-IR
Free-electron lasers	Any wavelength

PROBLEMS

2.1. (a) List three ways in which solid-state lasers may be pumped.
(b) Explain why it is necessary to have temperature control of the rod in solid-state lasers.

2.2. (a) What is a typical gas ratio by volume for a He-Ne laser? (e.g. He:Ne–XX:YY).
(b) What is a typical total gas pressure in a He-Ne laser?
(c) How many different transitions normally occur in a He-Ne laser?
(d) What types of mirrors would typically be used in a He-Ne laser cavity?
(e) Would you prefer the mirrors in a He-Ne laser to be internal or external to the discharge tube? Why?

2.3. How would you expect the gain of an Ar^+ ion laser to vary with the internal diameter of the discharge tube? Explain.

2.4. (a) What is the purpose of nitrogen in the CO_2 laser?
(b) What is the purpose of helium in the CO_2 laser?
(c) What is the purpose of CO_2 in the CO_2 laser?
(d) In CO_2 lasers, transition occurs at 10.6 and at 9.6 μm. What method would be used to ensure that only the 10.6 μm wavelength beam is available at the output?

(e) Which would you recommend for butt welding of relatively thick steel, slow axial flow or transverse flow CO_2 laser? Explain.

2.5. (a) What type of CO_2 laser would you recommend for cutting relatively thin materials?
(b) What type of CO_2 laser would you recommend for butt-welding two thin pieces?
(c) Everything else being the same, which laser type would you recommend for cutting a given material, CO_2 or Nd:YAG? Explain.

2.6. Give two reasons why excimer lasers are a serious candidate for materials processing.

2.7. Show why $p_v(E_v)[1 - p_c(E_c)]$ increases, and $p_c(E_c)[1 - p_v(E_v)]$ decreases as temperature increases.

2.8. (a) Relative to most lasers, the diode laser is highly efficient. Why?
(b) Give two reasons why double heterojunction diode lasers require a lower threshold current density than homojunction diode lasers.
(c) Show that the divergence of a semiconductor laser should be about 5–15°.

2.9. Determine the power efficiency of a semiconductor laser which is operated in pulsed mode if the power input to the laser is 20 W, and the input current pulse is rectangular in shape. The peak output power is 7.5 W, with a duration of 0.5 ps, and the pulse repetition rate is 10 kHz. Further determine the duty cycle of the laser. The duty cycle is the ratio of the time that the laser is on to the total time.

BIBLIOGRAPHY

Abil'siitov, G.A. and Velikhov, E.P. (1984). Application of CO_2 lasers in mechanical engineering technology in the USSR. *Optics and Laser Technology* 16 (1): 30–36.

Baird, B. (2018). Fiber laser marking and welding of polymers finds new applications. *Industrial Laser Solutions* 11–13.

Barrett, C.R., Nix, W.D., and Tetelman, A.S. (1973). *The Principles of Engineering Materials*. Englewood Cliffs, New Jersey: Prentice Hall.

Chiu, M.-H., Hsu, S.-N., and Yang, H. (2004). D-type fiber optic sensor used as a refractometer based on total-internal reflection heterodyne interferometry. *Sensors and Actuators B* 101: 322–327.

Duley, W.W. (1982). *Laser Processing and Analysis of Materials*. New York: Plenum Press.

Harry, J.E. (1974). *Industrial Lasers and their Applications*. London: McGraw Hill.

Hecht, J. (2012). Fiber lasers: the state of the art. *Laser Focus World* 48 (4): 1–14.

Henderson, R.A. (1997). *A Guide to Laser Safety*. London, U.K.: Chapman & Hall.

Koechner, W. (1995). *Solid-State Laser Engineering*, 4the. Berlin, Germany: Springer.

Laud, B.B. (1985). *Lasers and Non-Linear Optics*. New Delhi: Wiley Eastern Limited.

Limpert, J., Roser, F., Klingebiel, S. et al. (2007). The rising power of fiber lasers and amplifiers. *IEEE Journal of Selected Topics in Quantum Electronics* 13 (3): 537–545.

Liu, X., Du, D., and Mourou, G. (1997). Laser ablation and micromachining with ultrashort laser pulses. *IEEE Journal of Quantum Electronics* 33 (10): 1706–1716.

Loosen, P., Treusch, G., Haas, C.R. et al. (1995). High-power diode-lasers and their direct industrial applications. *SPIE* 2382: 78–88.

Luxon, J.T. and Parker, D.E. (1985). *Industial Lasers and Their Applications*. Englewood Cliffs, New Jersey: Prentice Hall.

Mann, K. and Morris, T. (2004). Disk lasers enable application advancements. *Photonics Spectra* 38 (1): 106–110.

Marabella, L.J. (1994). *March 1994*. Automotive Laser Applications Workshop, Dearborn, Michigan: Future solid state lasers.

Nilsson, J. and Payne, D.N. (2011). High-power fiber lasers. *Science* 332: 921.

O'Shea, D.C., Callen, W.R., and Rhodes, W.T. (1977). *Introduction to Lasers and Their Applications*. Reading, Massachusetts: Addison-Wesley.

Quintino, L., Costa, A., Mirand, R. et al. (2007). Welding with high power fiber lasers – A preliminary study. *Materials and Design* 28: 1231–1237.

Richardson, D.J., Nilsson, J., and Clarkson, W.A. (2010). High power fiber lasers: current status and future perspectives. *J. Opt. Soc. Am. B* 27 (11): B63–B92.

Rofin-Sinar. *Technical Note* - CO_2 *Laser Welding*, Rofin-Sinar, Plymouth, Michigan.

Shi, W., Fang, Q., Zhu, X. et al. (2014). Fiber lasers and their applications. *Applied Optics* 53 (28): 6554–6568.

Shimoda, K. (1986). *Introduction to Laser Physics*, 2e. Berlin: Springer-Verlag.

Solarz, R., Beach, R., and Hackel, L. (1994). The future of diode pumped solid state lasers and their applicability to the automotive industry, *Automotive Laser Applications Workshop*, Dearborn, Michigan.

Svelto, O. (2010). *Principles of Lasers*, 5e. LLC, New York: Springer Science+Business Media.

Thyagarajan, K. and Ghatak, A.K. (2010). *Lasers, Theory and Applications*, 2e. LLC, New York: Springer Science+Business Media.

Tiffany, W.B. (1985). Drilling, marking, and application for industrial Nd: YAG Lasers. *SPIE, Applications of High Powered Lasers* 527: 28–36.

Walsh, C.A., Bhadeshia, H.K.D.H., Lau, A. et al. (2003). Characteristics of high-power diode-laser welds for industrial assembly. *Journal of Laser Applications* 15 (2): 68–76.

Wilson, J. and Hawkes, J.F.B. (1987). *Lasers - Principles and Applications*. New York: Prentice Hall International Series in Optoelectronics.

3 Beam Delivery

Beam delivery is the means by which the laser beam propagates from the generator to the target of interest. This normally involves the use of mirrors, lenses, fiber optic cables, or any combination of these elements. Since light is part of the electromagnetic spectrum, the manner in which the beam propagates is determined by the characteristics of the electromagnetic spectrum. Thus, we start our discussion in this chapter with a brief introduction to the electromagnetic spectrum. This is followed by a brief review of the Brewster angle. Next, birefringence is discussed, followed by polarization, expanders, and beam splitters. A general discussion on beam delivery systems then follows. Fiber optic systems are discussed as a major component of the chapter. Finally, diffractive optics and coherent beam combining (CBC) with optical phase array (OPA) technique (or Civan method) are discussed as effective methods for beam shaping.

3.1 THE ELECTROMAGNETIC SPECTRUM

The electromagnetic spectrum, of which visible light is part, exhibits both wave-like and particle-like properties. The radiations of the electromagnetic spectrum are characterized by their wavelength λ and frequency ν, which are related by:

$$\nu\lambda = c \tag{3.1a}$$

where c is the velocity at which electromagnetic radiation propagates in free space and is given by:

$$c = \frac{1}{\sqrt{\mu_{m0}\varepsilon_{p0}}} \approx 3 \times 10^8 \ \text{m/s} \tag{3.1b}$$

The velocity, c_m, at which an electromagnetic wave propagates in any medium is given by:

$$c_m = \frac{1}{\sqrt{\mu_m \varepsilon_p}} \tag{3.1c}$$

where μ_{m0} and ε_{p0} are the magnetic permeability and electric permittivity, respectively, in free space, and μ_m and ε_p are those in the medium.

c and c_m are related by the refractive index of the medium, n, as:

$$n = \frac{c}{c_m} = \sqrt{\frac{\mu_m \varepsilon_p}{\mu_{m0}\varepsilon_{p0}}} = c\sqrt{\mu_m \varepsilon_p} \tag{3.1d}$$

Principles of Laser Materials Processing: Developments and Applications, Second Edition. Elijah Kannatey-Asibu, Jr.
© 2023 John Wiley & Sons, Inc. Published 2023 by John Wiley & Sons, Inc.
Companion Website: www.wiley.com/go/Kannatey-Asibu/PrinicplesLaserMaterialsProcessing

From Eq. (3.1a), it can be shown (Problem 3.9) that for small $\Delta\lambda$,

$$\Delta\nu = -\left(\frac{c}{\lambda^2}\right)\Delta\lambda \quad \text{or} \quad \Delta\lambda = -\left(\frac{c}{\nu^2}\right)\Delta\nu \tag{3.2}$$

The negative sign arises since the wavelength increases as the frequency decreases. Let us now look at some basic characteristics of electromagnetic radiation that are associated with their propagation.

3.2 BIREFRINGENCE

Birefringence is the phenomenon whereby a light beam that is incident on a material is refracted into two rays, the ordinary ray and the extraordinary ray. Materials that exhibit this characteristic are said to be birefringent (optically anisotropic). The two rays are plane-polarized (see next section) in orthogonal directions. The ordinary ray obeys Snell's law and has the same velocity in any direction in the material. The extraordinary ray, on the other hand, has different velocities in different directions and does not follow Snell's law. Both the ordinary and extraordinary rays have the same velocity along the optic axis of the material. Since the two rays generally have different velocities, the birefringent material normally has two different refractive indices. An example of such a material is the calcite crystal.

3.3 BREWSTER ANGLE

In Figure 1.9, the electric and magnetic vectors of the light wave are shown as lying in specific planes. Light having this characteristic where the electric vector is restricted to a specific plane is said to be *linearly* or *plane-polarized*. The electric vector for a plane-polarized beam is illustrated schematically in Figure 3.1a, while that for an unpolarized beam is illustrated in Figure 3.1b.

The electric vector associated with a beam that is incident on a surface can be resolved into two components, one parallel to the plane of incidence, *p-component*, and the other normal to the plane of incidence, *s-component*. The intensities of the two components are equal if the original beam is unpolarized. However, the intensities of the components of the reflected beam are not necessarily equal and depend on the reflection coefficients for the two components of the reflected beam, which are different. These are given by:

$$R_{\mathrm{p}} = \frac{\tan^2(\theta_1 - \theta_2)}{\tan^2(\theta_1 + \theta_2)} \tag{3.3a}$$

$$R_{\mathrm{s}} = \frac{\sin^2(\theta_1 - \theta_2)}{\sin^2(\theta_1 + \theta_2)} \tag{3.3b}$$

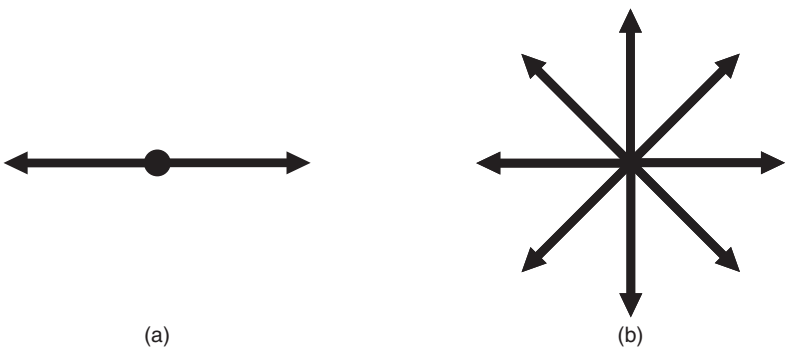


Figure 3.1 Electric vector of (a) a plane-polarized light and (b) an unpolarized light.

where R_p is the reflection coefficient of the reflected ray that is plane-polarized in the plane of incidence and R_s is the normal component. θ_1 and θ_2 are the respective incident and refracted angles. R_p and R_s are the intensity or power reflection coefficients and are also known as *reflectances*. The corresponding amplitude reflection coefficients are given by:

$$R_{pa} = \frac{\tan(\theta_1 - \theta_2)}{\tan(\theta_1 + \theta_2)} \tag{3.4a}$$

$$R_{sa} = \frac{\sin(\theta_1 - \theta_2)}{\sin(\theta_1 + \theta_2)} \tag{3.4b}$$

Variation of the reflection coefficients with the angle of incidence is shown in Figure 3.2a. From Eq. (3.3a), we observe that when $\theta_1 + \theta_2 = 90°$, the reflection coefficient, R_p, for the p-component in the plane of incidence is zero, since $\tan(\theta_1 + \theta_2)$ is then infinity. The reflected beam is then plane-polarized normal to the plane of incidence (Figure 3.2b). This occurs when the reflected and refracted rays are perpendicular to each other. The angle of incidence at which this occurs is referred to as the *polarizing angle* or *Brewster angle*, θ_B. Since $\theta_1 + \theta_2 = 90°$, we have $\sin\theta_2 = \cos\theta_1$, and thus from Snell's law, we have

$$\tan\theta_B = \frac{n_2}{n_1} = n_{21} \tag{3.5}$$

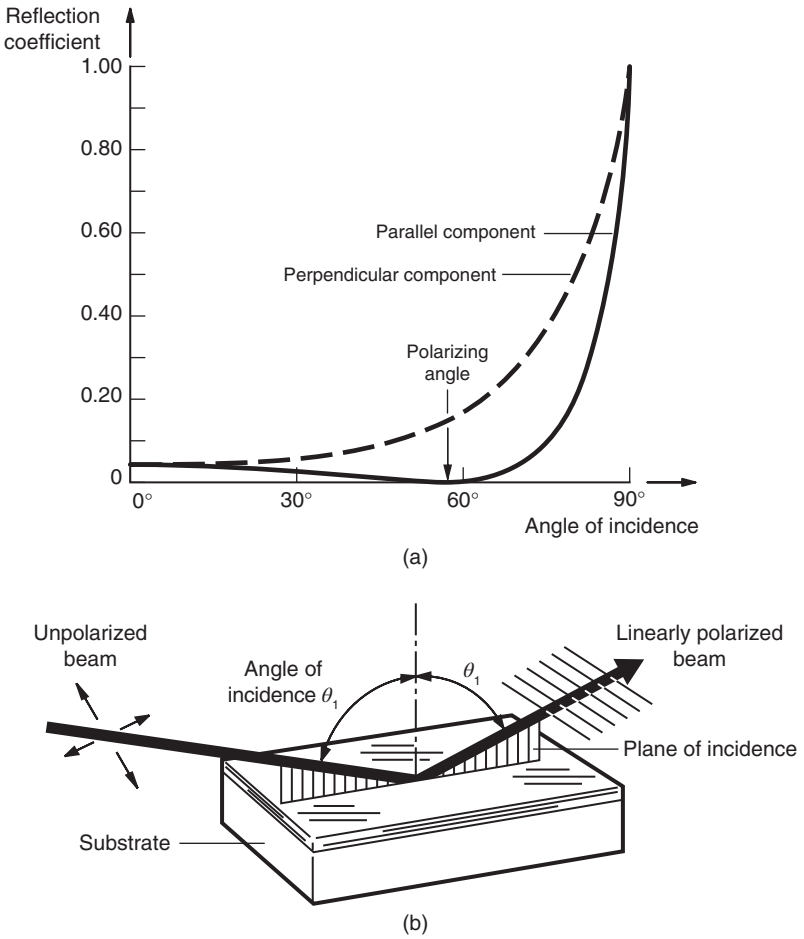


Figure 3.2 Reflection coefficients: (a) reflection coefficients for an air–glass interface ($n = 1.5$) and (b) reflection of an unpolarized beam at the polarizing angle. Source: From Dally and Riley (1991)/McGraw Hill.

When the original beam is incident from the medium with the lower refractive index, that is, medium 1, the beam undergoes a 180° (or $\lambda/2$ linear) phase shift during reflection. However, there is no phase shift if the light is incident from the higher refractive index medium.

The preceding discussion applies primarily to transparent materials such as glass. The reflection coefficients for metals are relatively large. The reflection coefficient for the s-component increases monotonously with increasing angle of incidence. However, that of the p-component first decreases with increasing angle of incidence until (depending on the material and wavelength of the incident beam), about 85°, when it begins to increase to 100% reflection at 90° incidence. Except for 0° incidence and 90° incidence, the reflection coefficient for the parallel component is always lower than that of the normal component.

3.4 POLARIZATION

In discussing polarization, we first consider a simplified one-dimensional propagation of the electric vector of an electromagnetic plane wave:

$$E_l(x,t) = E_0 \cos\frac{2\pi}{\lambda}(x - ct) = E_0 \cos(k_w x - \omega t) \tag{3.6}$$

where E_l is the magnitude of the electromagnetic wave (electric field vector), E_0 is the amplitude, $k_w = \frac{2\pi}{\lambda}$ is the wavenumber, and ω is the angular frequency. This is illustrated schematically in Figure 3.3a.

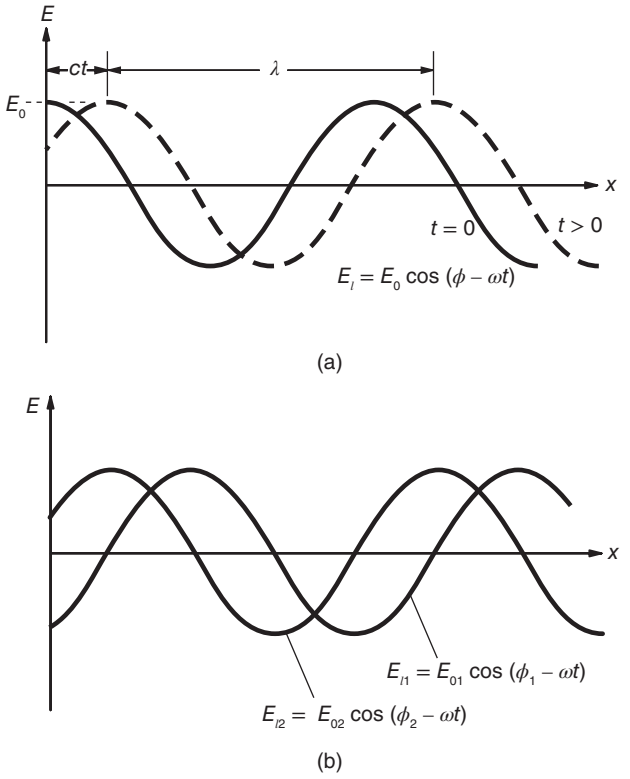


Figure 3.3 One-dimensional electromagnetic wave propagation: (a) a single wave at different times and (b) two different waves of the same frequency and a phase difference. Source: From Dallyand (1991)/McGraw Hill.

If we consider measurements that are made at a fixed position along the beam, then Eq. (3.6) can be expressed as:

$$E_l(t) = E_0 \cos(\varphi - \omega t) \tag{3.7}$$

where ϕ is the phase angle of the wave at the position of interest.

Two different waves, 1 and 2, of the same frequency (Figure 3.3b) can be represented as:

$$E_{l1}(t) = E_{01} \cos(\varphi_1 - \omega t) \tag{3.8}$$

and

$$E_{l2}(t) = E_{02} \cos(\varphi_2 - \omega t) \tag{3.9}$$

If these two waves are plane-polarized in the same plane and are superimposed on each other, then the resulting light vector $E_l(t)$ is given by:

$$E_l(t) = E_{l1}(t) + E_{l2}(t) \tag{3.10}$$

which can be shown to reduce to the form:

$$E_l(t) = E_0 \cos(\varphi - \omega t) \tag{3.11}$$

where

$$E_0^2 = E_{01}^2 + E_{02}^2 + 2E_{01}E_{02} \cos(\varphi_2 - \varphi_1) \tag{3.12}$$

and

$$\tan \varphi = \frac{E_{01} \sin \varphi_1 + E_{02} \sin \varphi_2}{E_{01} \cos \varphi_1 + E_{02} \cos \varphi_2} \tag{3.13}$$

Thus, the frequency of the resulting light vector is the same as that of the original light vectors. However, the amplitudes and phase angles are different.

Now let us consider the situation where the two plane waves $E_{ly}(t)$ and $E_{lz}(t)$ lie in different planes which are orthogonal to each other (Figure 3.4). The resulting light vector will then have a magnitude of

$$E_l(t)^2 = E_{ly}(t)^2 + E_{lz}(t)^2 \tag{3.14}$$

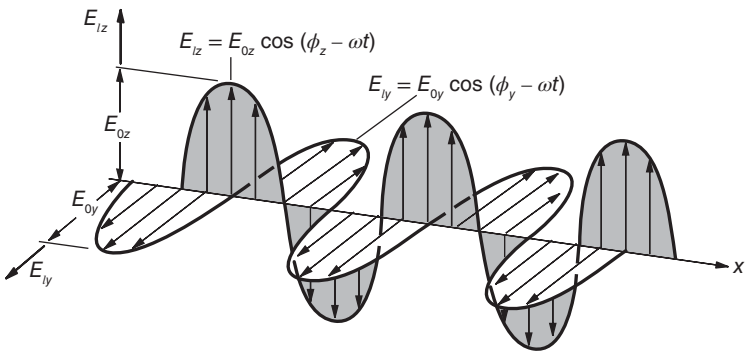


Figure 3.4 Plane-polarized electromagnetic waves in two orthogonal planes.

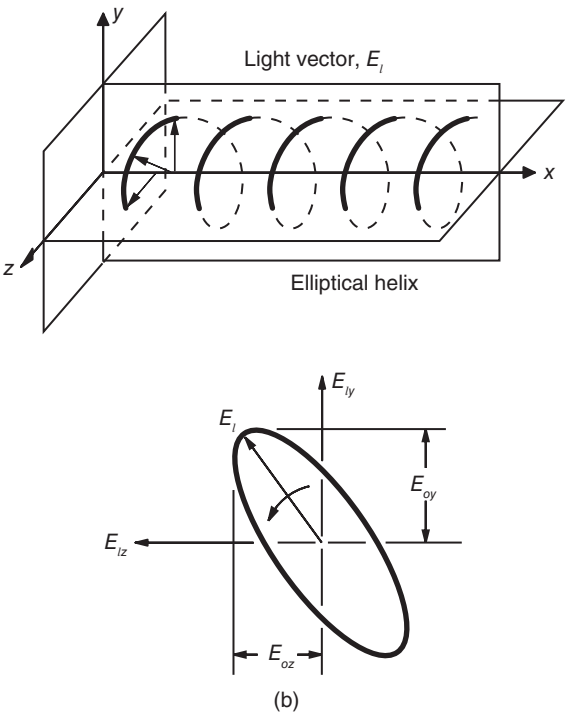


Figure 3.5 (a) Instantaneous configuration of the resultant electric vector of two plane-polarized electromagnetic waves in two orthogonal planes and (b) normal projection of the resultant electric vector of two plane-polarized electromagnetic waves in two orthogonal planes. Source: From Dally and Riley (1991)/McGraw Hill.

If we go back to the original equation for the wave motion, we can show that at any moment in time, the resultant electric vector along the x-axis or direction of propagation describes an elliptical helix (Figure 3.5a). If we now look at a projection of this resultant vector on a plane perpendicular to the x-axis, the equation describing the outermost trace of the projected image can be obtained by eliminating time from Eqs. (3.8) and (3.9), giving

$$\left[\frac{E_{ly}(t)}{E_{0y}}\right]^2 - 2\frac{E_{ly}(t)E_{lz}(t)}{E_{0y}E_{0z}}\cos(\varphi_z - \varphi_y) + \left[\frac{E_{lz}(t)}{E_{0z}}\right]^2 = \sin^2(\varphi_z - \varphi_y) \tag{3.15}$$

which is the equation of an ellipse (Figure 3.5b). Thus, the electromagnetic radiation obtained by superimposing two beams that are plane-polarized in two orthogonal planes is said to be *elliptically polarized*. The projected electric vector rotates clockwise in the projected plane as time varies, and the resulting helix is a right circular helix if the phase difference is such that

$$\varphi_z - \varphi_y = \frac{2m-1}{2}\pi \quad m = 0, 1, 2, 3, \ldots \tag{3.16}$$

and rotates in the counterclockwise direction, producing a left circular helix if

$$\varphi_z - \varphi_y = \frac{2m+1}{2}\pi \qquad m = 0, 1, 2, 3, \ldots \tag{3.17}$$

If $E_{0y} = E_{0z} = E_0$, that is, the amplitudes of the two light vectors are equal, and the phase difference between them is 90° or a quarter of a wavelength, then Eq. (3.15) becomes

$$E_{ly}(t)^2 + E_{lz}(t)^2 = {E_0}^2 \tag{3.18}$$

This is the equation of a circle. The resulting light is then said to be *circularly polarized*.

Plane-polarized light may be obtained by using a special optical element (plane polarizer) such as a Polaroid sheet, or by positioning a reflector such that the light beam is incident on it at the Brewster angle. Furthermore, if we compare Eqs. (3.3a) and (3.3b) for R_p and R_s, we find that R_s is always greater than R_p. Thus, repeated reflections from mirrors will eventually result in light that is plane-polarized normal to the plane of incidence.

3.5 BEAM EXPANDERS

Beam expanders are normally used for changing the size of the output beam. From Eq. (1.84), it is evident that the size of the focused beam is inversely proportional to the original beam size:

$$w_f = \frac{1}{\pi}\frac{f_l \lambda}{w_o}$$

Thus, a more intensely focused beam can be obtained from a laser beam of the same power if its diameter as it enters the focusing lens is increased. This could be achieved using a beam expander, which essentially consists of a series of mirrors (Figure 3.6a) or lenses (Figure 3.6b).

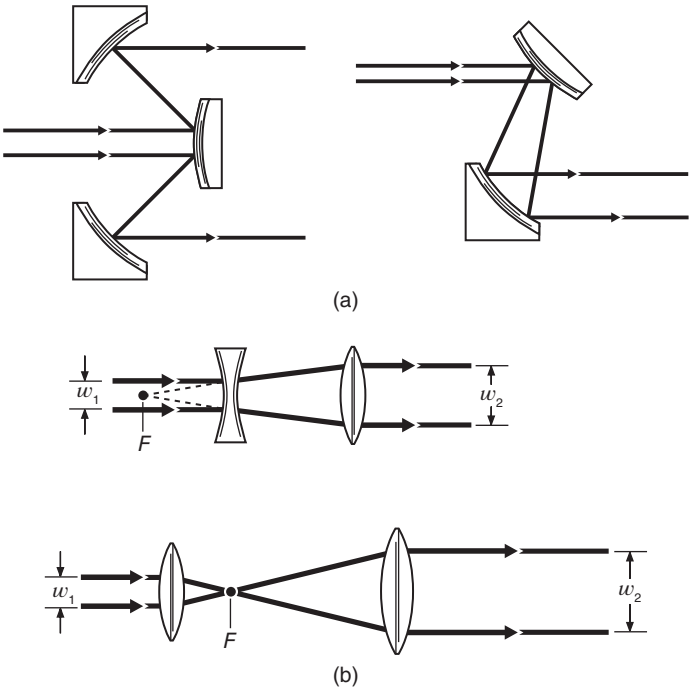


Figure 3.6 Examples of beam expanders: (a) mirrors and (b) lenses.

The transmissive (lens) beam expanders are mounted in such a way that they have the same focal point (Figure 3.6b). The size of the expanded beam, w_2, is given by:

$$w_2 = w_1 \frac{f_{l2}}{f_{l1}} \tag{3.19}$$

where w_1 is the size of the incoming beam, w_2 the size of the expanded beam, f_{l1} the focal length of the lens on which the original beam is incident, and f_{l2} the focal length of the output lens.

Increasing the beam size also has the advantage of reducing the diffraction spreading associated with it (Section 1.7.1).

For the lens system, another configuration is based on a non-focusing arrangement. This type has the advantage of being more compact and suitable for high-power applications where the focused beam could present problems. The focusing arrangement, on the other hand, enables the original beam to be focused at a pinhole where higher-order diffraction fringes are filtered or blocked out. This results in a beam whose intensity distribution is more uniform because of the noise reduction.

3.6 BEAM SPLITTERS

Beam splitters are optical elements that are used to divide an incoming beam into two, and through cascading, into any number of parts. One simple method is based on reflection, using a wedge-shaped reflector (Figure 3.7), with the edge pointing directly into the laser beam path to reflect the incoming beam into two parts. For the split beams to come out parallel, the wedge angle has to be 90°. The power ratio of the split beams is determined by the position of the wedge edge relative to the axis of the incoming beam. The quality of the beams depends on the dimensional accuracy of the wedge and smoothness of the reflector.

Another technique makes use of a partially transmissive and partially reflective optical element (Figure 3.8). When the laser beam is incident on such a material, part of the beam is reflected, while the remaining portion is transmitted, separating the beam into two parts. Such an optical element is normally obtained by coating a transparent material such as a *ZnSe* block with a reflective material such as inconel. The amount of coating determines the ratio of transmitted to reflected beam power. The latter can thus be varied for a given application by using an element with a graded or varied

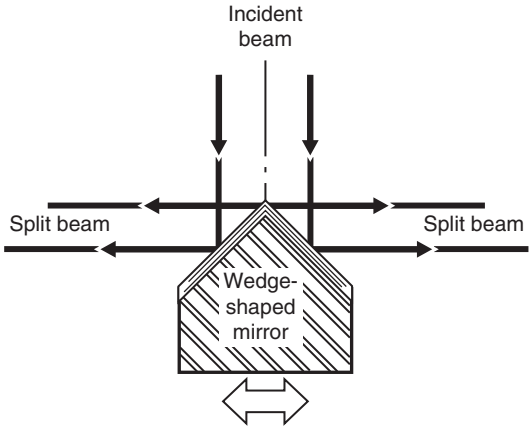


Figure 3.7 Wedge-shaped beam splitter. Source: From Chryssolouris (1991)/Springer Nature.

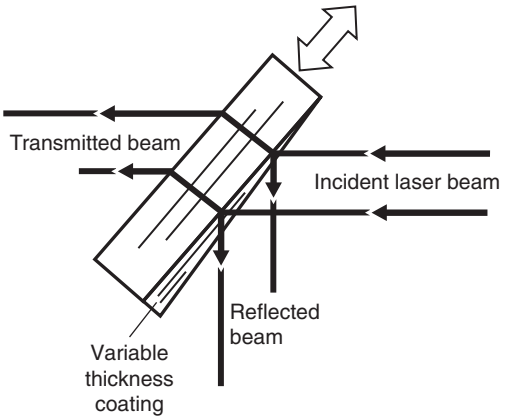


Figure 3.8 Partially reflective, partially transmissive beam splitter. Source: From Chryssolouris (1991)/McGraw Hill.

coating density on the surface and changing the location at which the beam is incident on the surface of the optical element.

Other techniques for splitting laser beams are based on the principles of the Fabry–Perot interferometer or a combined linear polarizer and electro-optic cell.

3.7 BEAM DELIVERY SYSTEMS

A beam delivery system is a means for directing the laser beam from the generator to the point of application. The two principal types are as follows:

1. Conventional beam delivery.
2. Fiber optic beam delivery.

3.7.1 Conventional Beam Delivery

Conventional beam delivery systems are conceptually simple, as shown schematically in Figure 3.9a, typically consisting of three main components:

1. The beam bending assembly, which is usually reflective, that is, consists of mirrors.
2. The focusing assembly, which may be transmissive or reflective.
3. Interconnecting beam guard tubes.

The beam bender assembly usually contains mirrors oriented at 45°, which reflect the unfocused beam at right angles to the original direction. Common materials used for the mirrors include copper or molybdenum, which may be bare or coated with silicon. The advantage of bare molybdenum is that it is durable and easily cleaned, making it suitable for industrial applications. However, it absorbs about 1–4% of the incident light. The coated version, on the other hand, absorbs only about 0.1% of the light, but cannot be restored by cleaning.

The lens is contained in the focusing assembly, which is normally made adjustable so that the distance between the focal point and nozzle tip can be varied to suit the application. It also has an inlet for shielding or cutting assist gas flow. For high-power applications, both the beam bender and

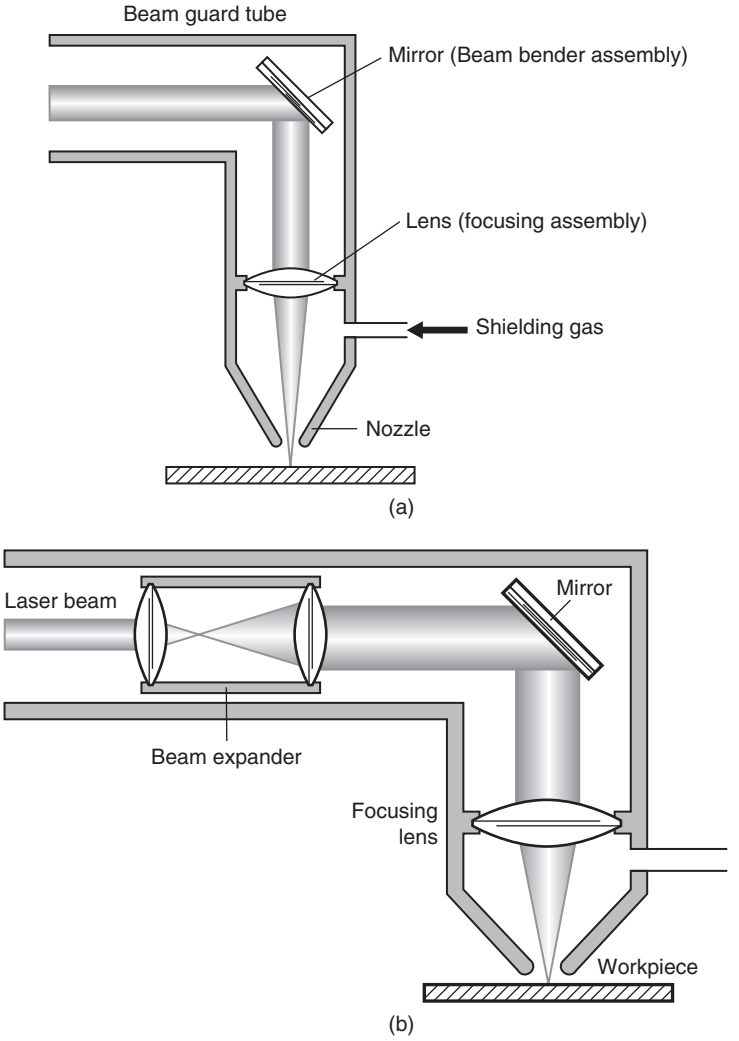


Figure 3.9 (a) Simplified laser beam delivery system and (b) beam delivery system with an expander.

focusing assembly have built-in circulating systems for air or water cooling to prevent overheating of the optics, which may cause thermal lensing or distortion of the laser beam, and thereby change the focusing characteristics of the lens. Common lens materials include *ZnSe, KCl, GaAs*, and *CdTe. ZnSe* is more commonly used with CO_2 lasers.

A number of beam benders can be used depending on the application. The workstation is sometimes placed a distance of about 6 m or more from the output coupler to permit the beam to be fully developed in the far-field mode and also to minimize the impact of back reflection into the laser generator. Immediately behind the output coupler (aperture), the resulting wavefront is not planar. This is due to diffraction effects that result from Huygen's principle. A planar wavefront is fully developed after a distance of about five Rayleigh ranges (Section 8.1.3).

Where necessary, a beam expander may be inserted in the beam path to provide the desired beam size (Figure 3.9b). For high-power lasers in the infrared range, it is normal practice to provide beam guards made from aluminum tubes for the beam along its path from the generator to the bender for

safety since the beam is invisible. The beam guard also serves to keep dust and other contaminants such as oil and mist off the delivery system. Internal cleanliness is assured by flowing clean dry air or nitrogen through the optical path in a direction away from the mirrors and lenses, at a pressure of about 30 psi.

Fiber optic systems provide a very flexible means of beam delivery. However, currently available systems are heavily absorbed by the longer wavelength CO_2 lasers and are thus suitable primarily for the shorter wavelength lasers, for example, the Nd:YAG laser. The basic principles of fiber optic systems will be discussed next.

3.7.2 Fiber Optic Systems

Optical fiber systems are used in transmitting laser beams or electromagnetic energy from one location to another. This application is quite common in optical communications where the power levels are generally low. For materials processing, where high-power industrial lasers are typically used, usage of optical fibers for transmitting the laser beam is limited to the shorter wavelength lasers. The longer wavelength CO_2 laser results in heating of the cable. Optical fibers are thus not used for high-power CO_2 laser beam transmission. The use of fiber optic transmission makes beam delivery more flexible, compared to the conventional mirrors and lenses.

The basic structure of an optical fiber system is shown schematically in Figure 3.10. The principal components include:

Transmitter: Which converts an electrical signal into an optical signal. This includes a light source.

Transmission Line: Which conducts the optical signal, and this comprises the optical fibers.

Receiver: Which converts the optical signal back to the original electrical form. The receiver normally consists of photodetectors and signal amplifiers.

Our focus in this section will be on the transmission line since fiber optic cables are used in laser processing primarily to transmit a beam from the source to the delivery point without any electrical conversions. First, we look at the characteristics of optical fibers in general.

3.7.2.1 Optical Fiber Characteristics Optical fibers can be characterized on the basis of either their physical or optical properties. On the physical side, optical fibers consist of glassy fibers of diameter about 125 μm each, with lengths up to the order of kilometers. Inorganic oxide glasses of high silica content are often used, and the resulting fibers are very flexible. For the optical characteristics, the major limitations of optical fibers are the attenuation and bandwidth dispersion. Attenuation results in dissipation of optical energy during transmission and is caused mainly by absorption and scattering. The bandwidth determines the limits of frequencies that can be transmitted.

Optical fibers are normally operated in the temperature range from −55 to 125 °C but may extend to −250 to 500 °C. Due to the dielectric nature of glass and the small size of fibers (that is, much smaller than the wavelength of radio and microwaves), they are immune to interferences from such

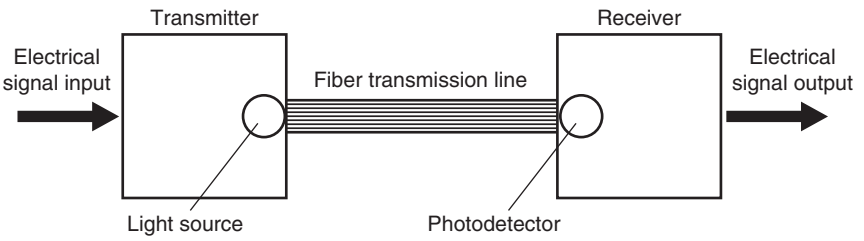


Figure 3.10 Schematic of a fiber optic system.

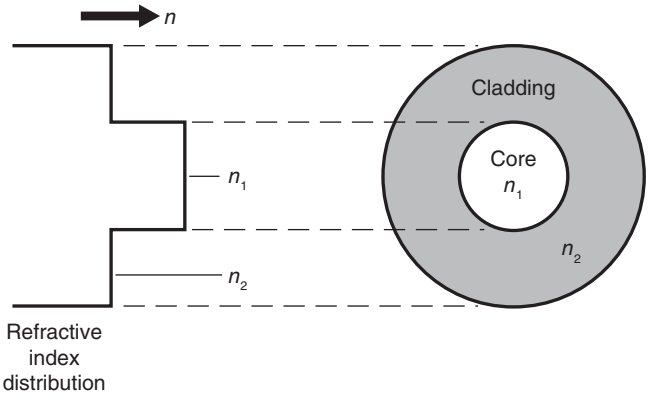


Figure 3.11 Basic structure of a fiber optic cable.

signals. Thus, fibers can pass through certain regions of high electromagnetic fields without interference. Also, waveguide covers prevent interference from other optical signals.

3.7.2.2 Waveguide Structure The optical fiber waveguide is the thread-like medium through which the light propagates. It is generally circular in section and has a long cylindrical structure. The basic structure has two components (Figure 3.11):

Core: This is the inner portion through which light is transmitted and is made of transparent glass (or plastic) of relatively high refractive index. Core diameters typically vary between 4 and 12 μm for single-mode fibers (see Section *Single-Mode Fibers*), and between 50 and 200 μm for multimode fibers (see Section *Step-Index Multimode Fibers*).

Cladding: This is the concentric outer layer with a relatively lower index of refraction. It reduces scattering loss from discontinuities at the core surface, adds mechanical strength to the fiber, and protects the core from absorbing surface contaminants. The cladding may also be a glass or plastic material.

The difference in refractive indices of the core (n_1) and cladding (n_2) is often expressed in the normalized form, Δ, as:

$$\Delta = \frac{n_1 - n_2}{n_1} \tag{3.20}$$

Typical values for Δ would be 0.2% for single-mode and 1% for multimode fibers. Since the core has a higher index than the cladding, total internal reflection occurs, enabling the waves to stay within the fiber. The optical fiber waveguide is itself contained in another coaxial layer that supports the fiber structure.

3.7.2.3 Background Consider the simple system shown in Figure 3.12, with a core of radius, r_c. Let the light incident on the fiber end make an angle θ with the waveguide axis. If θ is less than θ_a, the *angle of acceptance*, total internal reflection occurs because the refracted light is then incident at the core–cladding interface at an angle greater than θ_c, the critical angle. The light then stays within the core. For angles greater than θ_a, no internal reflection will occur and the light will propagate into the cladding. The incident rays that propagate within the core are more commonly defined using the *numerical aperture, NA*. It is a measure of the light-gathering or collecting power of an optical fiber and is defined as:

$$NA = \sin\theta_a \tag{3.21}$$

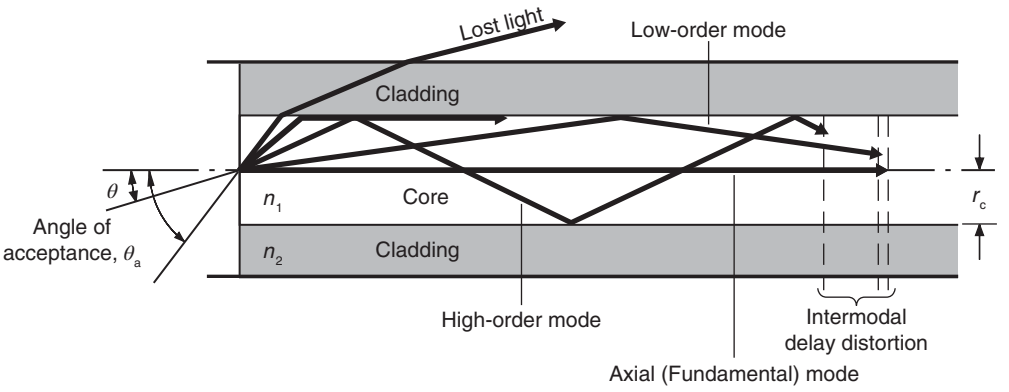


Figure 3.12 Propagation of a light ray in an optical fiber.

For step-index fibers, this is given by:

$$NA = \sqrt{{n_1}^2 - {n_2}^2}$$

The amount of light that can enter and stay within the fiber increases as the numerical aperture increases. However, the bandwidth that can be transmitted decreases as the NA increases.

Only specific quantized wave modes are propagated, and the mode volume parameter or normalized frequency, ν_n, of the modes guided by the fiber is defined as:

$$\nu_n = k_w r_c \sqrt{{n_1}^2 - {n_2}^2} = k_w r_c NA \tag{3.22}$$

where $k_w = 2\pi/\lambda$ is the wavenumber and r_c is the core radius.

The total number of modes, M_c, that can exist in a step-index fiber is approximated by:

$$M_c = \frac{4\nu_n^2}{\pi^2} \approx \frac{\nu_n^2}{2} \tag{3.23}$$

3.7.2.4 Fiber Types The three most common fiber types are as follows:

1. Step-index multimode fibers.
2. Single-mode fibers.
3. Graded-index multimode fibers.

Step-Index Multimode Fibers A step-index fiber is one with a uniform core refractive index, n_1, which abruptly changes at the core–cladding interface to the refractive index of the cladding, n_2 (Figure 3.11):

$$n_1 > n_2 \tag{3.24}$$

The core may be silica, with a refractive index of about 1.46. The step-index multimode fiber usually has a core cross-sectional area that is large enough for transmitting a significant amount of energy. Thus, the normalized frequency is high enough to support a number of discrete modes, that is,

$$V_n \geq 2.405$$

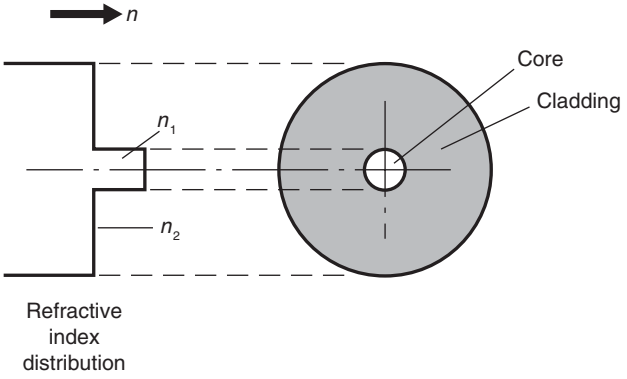


Figure 3.13 Cross section of a single-mode fiber.

It has the disadvantage that the low-order modes (with angle of incidence at the fiber end, θ, very small) travel a shorter ray path through the guide than do higher-order modes, that is, those with θ close to θ_a (Figure 3.12). This causes *intermodal dispersion* (Section 3.7.2.5) where rays starting at the same time become out of phase. It restricts the fiber bandwidth or the frequency range over which it is useful. For this and other reasons, the step-index multimode fiber is only suitable for short-distance applications.

Single-Mode Fibers Single-mode fibers are another form of step-index fibers. However, the core is much smaller. Also, the difference in refractive index between the core and cladding is much smaller. As the normalized frequency, ν_n, is reduced, so does the number of propagated modes. When $\nu_n < 2.405$, all the higher-order modes are cut off, and only the fundamental mode propagates. This reduction in ν_n is achieved either by a reduction in core diameter or a reduction in the index difference, as shown in Eq. (3.22). All modes that are incident at an angle are therefore cut off, and propagation occurs only along the fiber axis. This eliminates intermodal dispersion. A much greater bandwidth is thus obtained. However, due to the minute core size, installation is a lot more difficult, especially with respect to alignment during joining. To further minimize dispersion, the single-mode fiber requires a laser beam source. A single-mode fiber is illustrated schematically in Figure 3.13.

Graded-Index Fibers In graded-index fibers, the refractive index of the core varies (is graded) from a maximum, n_1, at the center of the core to the value n_2, of the cladding (Figure 3.14), according to the relationship:

$$n_r = \begin{cases} n_1 \sqrt{\left[1 - 2\Delta\left(\frac{r}{r_c}\right)^{k_1}\right]} & 0 \le r \le r_c \\ n_2 & r > r_c \end{cases} \tag{3.25}$$

where k_1 is a constant and n_r is the core refractive index at radius r.

$k_1 = \infty$ for step-index fibers. For $k_1 = 2$, light from a point source experiences periodic focusing in the fiber (Figure 3.14b). For modes with larger entrance angles θ, the travel path is longer. However, they also travel in a region where the refractive index is lower, and thus the wave velocity is greater. Thus, they reach a given point about the same time as the lower-order modes which travel mainly in the slower regions. This reduces the intermodal dispersion. Thus, the bandwidth increases accordingly. The graded fiber index is thus good for large bandwidth and medium-distance applications.

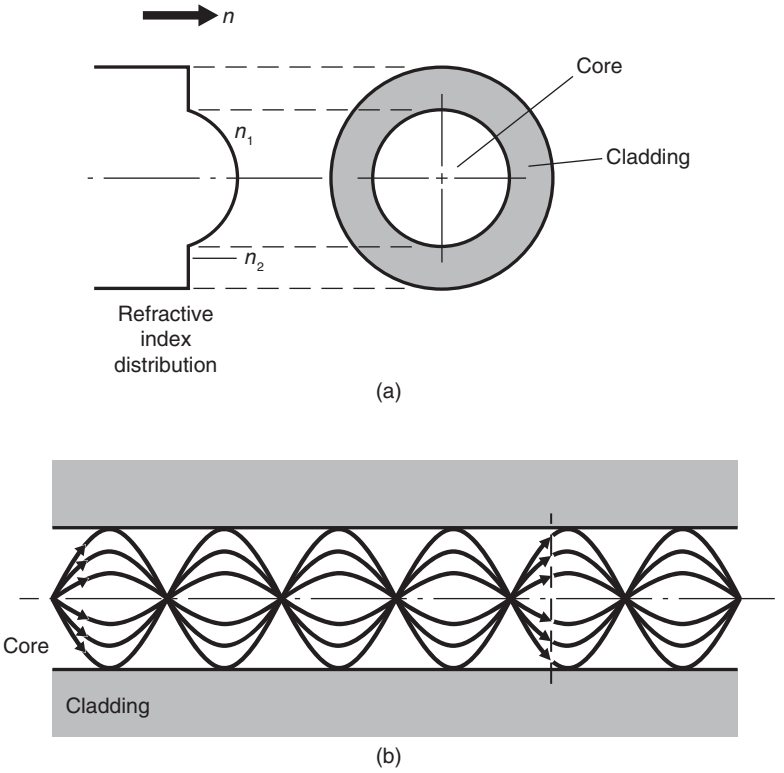


Figure 3.14 (a) Cross section of a multimode graded-index fiber and (b) periodic focusing in a multimode graded-index fiber. Source: From Cherin (1983)/McGraw Hill.

3.7.2.5 Beam Degradation As it propagates through a fiber, the light beam becomes degraded as a result of several factors that affect the beam characteristics. The two principal forms of degradation are as follows:

1. Dispersion or distortion.
2. Optical losses.

Before further discussing the degradation phenomena, we outline other ways in which optical fibers are characterized. The characteristics of an optical fiber can also be described in terms of its bandwidth or its impulse response in the time domain (rms pulse duration). Either of these characterizations can be obtained by exciting the fiber with an impulse input and observing the output pulse or its spectrum. The fiber characteristics are generally specified as either a bandwidth-length product, with units of megahertz kilometers, or a pulse dispersion denoted in nanoseconds per meter. We now take a look at the different forms of degradation.

Dispersion Dispersion or distortion of the light beam results from various causes. The principal contributing mechanisms are as follows:

1. Chromatic (material) dispersion.
2. Waveguide dispersion.
3. Modal dispersion.

To understand the phenomenon of dispersion, let us consider the concept of group velocity or group delay associated with the energy that is transmitted through the fiber. From Eq. (3.6), we have the following one-dimensional equation for the electric vector of a monochromatic electromagnetic plane wave propagating in a dielectric medium:

$$E_l(x,t) = E_0 \cos(k_{\mathrm{w}} x - \omega t) \tag{3.26}$$

By considering the phase of the wave to be constant, that is, $k_w x - \omega t =$ constant, we obtain the phase velocity (also known as wave velocity), u_{w}, as:

$$u_{\mathrm{w}} = \frac{dx}{dt} = \frac{\omega}{k_{\mathrm{w}}} \tag{3.27}$$

The phase velocity is also given, Eq. (3.1d), by:

$$u_{\mathrm{w}} = \frac{c}{n} \tag{3.28}$$

where n is the refractive index of the medium and c is light velocity in free space. Thus,

$$u_{\mathrm{w}} = \frac{c}{n} = \frac{\varpi}{k_{\mathrm{w}}} \tag{3.29}$$

For a nonmonochromatic source, the phase velocity in a dispersive medium may be determined using the center wavelength of the source. However, it is customary to use the group velocity, u_{g}, which in the limit, is defined as:

$$u_{\mathrm{g}} = \frac{dx}{dt} = \frac{d\omega}{dk_{\mathrm{w}}} \tag{3.30}$$

The group velocity is the velocity of the wave envelope and is the velocity at which information that is modulated on the wave will propagate. Generally, the group and phase velocities of a wave in a dispersive medium will be different. The inverse of the group velocity is referred to as the group delay per unit length, χ_{g}:

$$\chi_{\mathrm{g}} = \frac{1}{u_{\mathrm{g}}} = \frac{dk_{\mathrm{w}}}{d\omega} \tag{3.31}$$

Now from Eq. (3.29), this can also be expressed as:

$$\chi_{\mathrm{g}} = \frac{dk_{\mathrm{w}}}{d\omega} = \frac{1}{c}\frac{d(n\omega)}{d\omega} = \frac{1}{c}\left(n + \omega\frac{dn}{d\omega}\right) \tag{3.32}$$

The expression in brackets on the right-hand side of Eq. (3.32) is referred to as the group index, n_{g}:

$$n_{\mathrm{g}} = n + \omega\frac{dn}{d\omega} \tag{3.33}$$

Eq. (3.33) can also be expressed (Problem 3.5) as:

$$n_{\mathrm{g}} = n - \lambda\frac{dn}{d\lambda} \tag{3.34}$$

The group velocity and group delay can now be expressed as:

$$u_g = \frac{c}{n_g} \tag{3.35}$$

and

$$\chi_g = \frac{n_g}{c} \tag{3.36}$$

Chromatic Dispersion Chromatic dispersion is also sometimes referred to as material dispersion and results from the nonlinear variation of the refractive index of a material with wavelength, that is, change in n_1 and n_2 with λ. As discussed in Sections 1.5 and 1.7, even laser beams are not exactly monochromatic and generally cover a range of wavelengths. Thus, if the refractive index–wavelength relationship is not linear over that range, the different wavelengths will propagate with different velocities within the fiber, resulting in distortion of the optical signal.

Now consider a nonmonochromatic source with a wavelength $\Delta\lambda$ that spreads between λ_1 and λ_2 and is centered at λ_0. If it propagates energy over a single-mode fiber of length L_f, the difference in arrival times of energies propagated by the wavelengths λ_1 and λ_2, the delay distortion or pulse broadening, $\Delta\tau_{cd}$, can be expressed as:

$$\Delta\tau_{cd} = L_f[\chi_g(\lambda_1) - \chi_g(\lambda_2)] = \frac{L_f}{c} n_g(\lambda_1) - \frac{L_f}{c} n_g(\lambda_2) = -\frac{L_f}{c}\frac{dn_g}{d\lambda}\Delta\lambda \tag{3.37}$$

But from Eq. (3.34),

$$\frac{dn_g}{d\lambda} = \frac{dn}{d\lambda} - \frac{dn}{d\lambda} - \lambda\frac{d^2n}{d\lambda^2} = -\lambda\frac{d^2n}{d\lambda^2} \tag{3.38}$$

Substituting Eq. (3.38) into (3.37) gives the pulse broadening due to chromatic dispersion in a single-mode fiber as:

$$\Delta\tau_{cd} = \frac{L_f}{c}\lambda\frac{d^2n}{d\lambda^2}\Delta\lambda = L_f D_d \Delta\lambda \tag{3.39}$$

where D_d is the dispersion coefficient, with units ps/(km nm). This is shown schematically in Figure 3.15.

Example 3.1 Consider a beam of wavelength 1.0 μm propagating through a fiber of length 5 km and whose dispersion coefficient is as shown in Figure 3.15. If the spectral width, $\Delta\lambda$, of the beam is 0.050 μm, determine the pulse broadening due to chromatic dispersion.

Solution:
From Figure 3.15, the absolute value of the dispersion coefficient is obtained as 42 ps/(km nm).

Thus, the pulse broadening is

$$\Delta\tau_{cd} = L_f D_d \Delta\lambda = 5 \times 42 \times 50$$
$$= 10.5 \text{ ns}$$

The effect of chromatic dispersion can be minimized by appropriate selection of beam wavelength to fall in the region where $d^2n/d\lambda^2 \approx 0$. This is normally in the range 1.2–1.4 μm for fused silica. The impact of chromatic dispersion in multimode fibers is generally negligible. However, it is the more dominant form of distortion and thus limits the bandwidth in single-mode fibers.

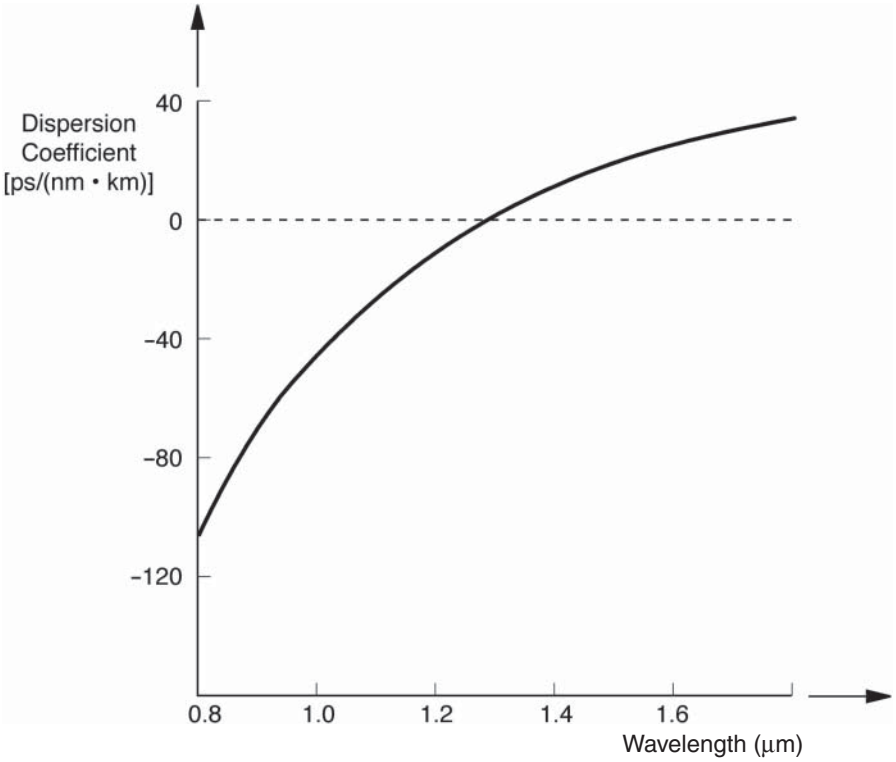


Figure 3.15 Dispersion coefficient as a function of wavelength. Source: From Keiser (1991)/McGraw Hill.

WAVEGUIDE DISPERSION Waveguide dispersion results from the fact that the propagating characteristics of a mode depend on the beam wavelength. Longer wavelengths have a longer path since they reflect at more oblique angles with the cladding. Analysis of the delay distortion for waveguide dispersion is beyond the scope of this book. It can be shown, however, that the pulse broadening, $\Delta\tau_{\mathrm{wd}}$, due to waveguide dispersion may be expressed as:

$$\Delta\tau_{\mathrm{wd}} = \frac{L_{\mathrm{f}}}{c}\frac{\Delta\lambda}{\lambda}(n_2 - n_1)D_{\mathrm{w}}(\nu_{\mathrm{n}}) \tag{3.40}$$

where $D_{\mathrm{w}}(\nu_{\mathrm{n}})$ is a dimensionless dispersion coefficient and L_{f} is the fiber length.

The pulse broadening due to waveguide dispersion is relatively small compared to chromatic dispersion.

INTERMODAL DISPERSION Intermodal dispersion results from differences in the distances propagated by the different modes sustained by the fiber (Figure 3.12). If all the rays shown in the figure start out at the same instant, the bouncing rays reach the end of the fiber at a later time than the axial ray. The temporal delay (dispersion) in the arrival times of the rays causes delay distortion or change in the spectrum of the original input beam. The derivation of the expression for the delay distortion for this case is also beyond the scope of this book. However, for a step-index multimode fiber, the group delay, χ_{g}, can be expressed as:

$$\chi_{\mathrm{g}} = \frac{L_{\mathrm{f}}}{c}\frac{d(n_2 k_{\mathrm{w0}})}{dk_{\mathrm{w0}}} + \frac{L_{\mathrm{f}}}{c}\frac{\nu_{\mathrm{n}}}{k_{\mathrm{w0}}}\frac{d(n_2 k_{\mathrm{w0}} b_{\mathrm{c}}\Delta)}{d\nu_{\mathrm{n}}} \tag{3.41}$$

where $k_{w0} = \frac{2\pi}{\lambda_0}$ is the wavenumber in free space, ν_n, n_2, Δ, and r_c are as defined in Sections 3.7.2.2 and 3.7.2.3, and b_c is a normalized propagation constant defined as:

$$b_c = \frac{(\gamma_c r_c)^2}{\nu_n^2} \tag{3.42}$$

where γ_c is the rate of decay of the electric field in the cladding.

The first term on the right-hand side of Eq. (3.41) corresponds to the group delay due to chromatic dispersion in the step-index multimode and is independent of a particular mode.

The second term corresponds to the modal waveguide delay. It is different for every mode. When a laser pulse enters the fiber, it is shared by many guided modes in the fiber. The pulse thus splits up into a number of pulses. Each of these exits the fiber at a different time since the delays of the modes are different. The difference in arrival times of the mode with the largest waveguide group delay and that with the least delay can be derived, through rigorous analysis (which is beyond the scope of this book) to be

$$\Delta\tau_m = \frac{L_f}{c}(n_1 - n_2)\left(1 - \frac{\pi}{\nu_n}\right) \tag{3.43a}$$

Equation (3.43a) expresses the difference in the arrival times of the leading and trailing edge of the output pulse of the fiber. A simpler approximation of the pulse broadening due to intermodal distortion may be based on the difference between the travel time of the slowest mode (corresponding to the highest-order mode) and that of the fastest (corresponding to the fundamental mode) (Figure 3.12). Using ray tracing, this is obtained as:

$$\Delta\tau_m = t_{max} - t_{min} = \frac{L_f n_1 \Delta}{c} \tag{3.43b}$$

where t_{max} is the travel time of the slowest mode (one with the longest ray congruence path) and t_{min} is the travel time of the fastest mode (one with the shortest ray congruence path).

Example 3.2 Determine the pulse broadening for a 5 km fiber with core and cladding refractive indices of $n_1 = 1.48$ and $n_2 = 1.46$, respectively.

Solution:
The approximate value of the pulse broadening is obtained as:

$$\Delta\tau_m = \frac{L_f n_1 \Delta}{c} = \frac{5000 \times 1.48}{3 \times 10^8}\left(\frac{1.48 - 1.46}{1.48}\right) = 333 \text{ ns}$$

Thus, the time delay between the axial ray and the ray that just enters the fiber at the angle of acceptance is about 333 ns. This corresponds to a modal dispersion-limited bandwidth of approximately 3.0 MHz.

It is evident that modal dispersion will generally not be a problem with single-mode fibers, since single-mode fibers only sustain a single mode of the beam. However, for step-index multimode fibers, modal dispersion is much greater than chromatic dispersion. It is the dominant form of distortion and is the primary mechanism that limits step-index multimode fibers to low bandwidth (less than 100 MHz km) applications.

Optical Losses There are several factors that contribute to optical losses associated with beam transmission by a fiber optic system. These can be categorized as:

1. Input-coupling losses.
2. Connector/splice losses.
3. Fiber losses.
4. Output-coupling losses.

The most significant of these losses, especially in regard to fiber optic beam delivery for manufacturing applications, are the input-coupling and fiber losses. We will thus provide further discussion only on these two sources of optical losses.

INPUT-COUPLING LOSSES Input-coupling losses occur at the interface between the beam source and the fiber. These include

1. Mismatch between the source's emitting area and the fiber core area, which contributes to input-coupling losses.
2. Effect of the numerical aperture, which defines the light-gathering ability of the fibers themselves.
3. The packing fraction, which is associated with fiber bundle cables, where several fibers are grouped together and illuminated by one light source. This loss is given by the ratio of the collective core areas of the fibers to the total bundle cross-sectional area.

FIBER LOSSES Losses in the fiber result in attenuation of the beam as it propagates through the fiber. The attenuation is expressed as the ratio of input power to output power (in decibels) per unit length of fiber:

$$\alpha_{\mathrm{l}} = \frac{1}{L_{\mathrm{f}}} 10 \log_{10} \left(\frac{q_{\mathrm{in}}}{q_{\mathrm{out}}} \right) \tag{3.44}$$

where α_{l} is attenuation, q_{in} is the input power, and q_{out} is the output power.

Fiber losses may be categorized into intrinsic and extrinsic losses.

Intrinsic Fiber Losses Intrinsic fiber losses are associated with the basic microstructural characteristics of the material used in making the fiber. They can generally be attributed to three main causes:

1. *Ultraviolet Absorption*: Results from excitation of electrons of the glass material from the valence to the conduction band by photons of ultraviolet wavelength and is thus electronic in origin. The absorption is wavelength-dependent and may be expressed as:

$$\alpha_{\mathrm{uv}} = K_{\mathrm{uv}} e^{-\frac{\alpha_{\mathrm{uv}}}{\lambda}} \ \mathrm{dB/km} \tag{3.45}$$

where K_{uv} is a constant which is a function of the dopant in the glass, α_{uv} is a constant (with units of microns), and the wavelength λ is also expressed in microns. The absorption peaks at approximately 0.14 μm wavelength for fused silica.

2. *Infrared Absorption*: This is due to the lattice vibrational modes of the fused silica and dopants. Depending on the type of dopant, the absorption peak for this type of loss may fall between 7 and 11 μm wavelengths. The power loss due to infrared absorption increases exponentially with wavelength and is expressed as:

$$\alpha_{\text{ir}} = K_{\text{ir}} e^{-\frac{\alpha_{\text{ir}}}{\lambda}} \text{ dB/km} \tag{3.46}$$

where K_{ir} and α_{ir} are constants, and the wavelength λ is expressed in microns.

3. *Rayleigh Scattering*: This results from inhomogeneities (such as density and composition variations) within the fiber due to the amorphous nature of glass. The microscopic density and composition variations cause local variations of the refractive index, which are of dimensions much smaller than a wavelength, and thus act as scattering centers. The Rayleigh scattering loss is strongly wavelength-dependent and is given by:

$$\alpha_{\text{Rs}} = \frac{K_{\text{Rs}}}{\lambda^4} \text{ dB/km} \tag{3.47}$$

where K_{Rs} is the Rayleigh scattering coefficient, with units (dB/km) μm^4.

The general dependence of the intrinsic losses on wavelength is illustrated in Figure 3.16 which shows an increase in the ultraviolet and Rayleigh scattering losses as the wavelength decreases, while the infrared losses increase with increasing wavelength. The end effect is that the total intrinsic loss curve has a minimum, usually in the near-infrared region between 0.6 and 1.6 μm, depending on the fiber material and dopants.

Extrinsic Fiber Losses Extrinsic fiber losses include impurity and structural imperfection losses, macrobending losses, and microbending losses.

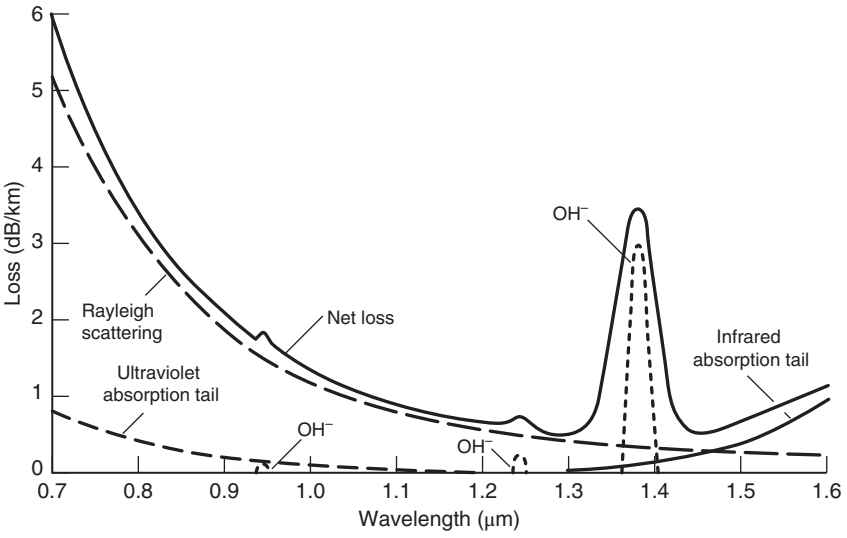


Figure 3.16 Sources of beam loss within a glass fiber. Source: From Cherin (1983)/McGraw Hill.

1. *Impurity and Structural Imperfection Losses*: These result from imperfections associated with the fabrication process, such as impurities within the glass and structural imperfections, and can be minimized by improvements to the fabrication process. An example of an impurity loss is illustrated by the peak at about 1.375 μm of the loss curve in Figure 3.16, caused by the hydroxyl group,OH^-.
2. *Macrobending Losses*: These are associated with axial bends of relatively large radius, of the order of 1 mm or more, in the fiber. This type of loss increases with increasing wavelength and decreasing bend radius. The loss is also mode-dependent, with higher-order modes being more severely attenuated by bending. Generally, bend radii of the order of a few centimeters should be avoided.
3. *Microbending Losses*: These are caused by random deflections of the fiber along its axis. The deflections are of the order of a few microns and can cause scattering losses.

3.7.2.6 Application of Optical Fibers in High-Power Laser Systems From the viewpoint of manufacturing applications, the principal effects that beam degradation (dispersion and attenuation) have on a laser beam as it propagates through an optical fiber are changes in the beam mode and intensity. Thus, a beam that enters the fiber with a Gaussian mode may be incident at the workpiece with a significantly different mode form and with the intensity significantly reduced. It is also obvious from Figure 3.16 that there is a certain wavelength range (typically in the near-infrared range) over which beam attenuation is minimized. Thus, it is more feasible to transmit the 1.06 μm wavelength Nd:YAG laser beam using fiberoptic cables than it is to transmit the 10.6 μm wavelength CO_2 laser beam, or the much shorter wavelengths in the ultraviolet region.

Since most high-power Nd:YAG lasers have beam quality of M^2 between 50 and 100, fiber sizes that are used typically range between 800 and 1000 μm in diameter. Smaller fiber sizes can only be used when beam quality significantly improves. This is because the beam size needs to be smaller than the fiber core diameter. Otherwise, power losses and heating may create significant problems.

Based on the preceding discussions, following are a few points to be noted in using fiber optic delivery systems in high-power laser applications.

1. Shorter fibers tend to have less effect on the beam characteristics, whereas a relatively long fiber may result in an output beam whose characteristics are significantly different from that of the incident beam.
2. Increasing the number of bends in the fiber cable or the tightness of the bends tends to generate more modes and thus change the beam profile or quality.
3. Smaller fibers limit the number of modes and thereby increase the beam quality.
4. Graded-index fibers tend to produce higher-quality beams.
5. Smaller fibers and graded-index fibers are more appropriate for processing applications that require high-power densities such as laser cutting.
6. For applications such as heat treatment that require a widely distributed beam (top-hat profiles), step-index fibers are preferable.
7. Fiber optic connectors are required at both the input end to launch the beam into the cable, and at the output end for delivering the beam to the workpiece.
8. The output optics system may also include a pair of lenses to collimate and focus the beam.

3.8 BEAM SHAPING

The mode (profile or shape) of a laser beam used in materials processing has significant impact on the process outcome (Section 8.1.3), influencing the melt pool shape and/or size, process window,

surface quality, and mitigating the formation of hot cracks. Changing the beam profile is thus an effective means of influencing the process dynamics. The shape of a beam is characterized by the irradiance profile, while the phase is defined by the optical path distance traveled by the wavefront. Beam shaping thus involves modulation of the irradiance and/or phase of a beam, coherent or incoherent.

Basic techniques that are used for transforming a beam include the following:

1. A small aperture to suppress the higher-order modes, leaving only the TEM_{00} mode (Section 1.2.2.3). A major drawback with this approach is the loss of a significant portion of the laser power. Furthermore, the transmitted beam is not necessarily uniform, but rather Gaussian.
2. Beam integrators (Section 11.1.2.2).
3. Fast beam deflection (wobbling or beam oscillation) (Section 10.7.3).
4. Field mapping transformation using
   i. Refractive optics.
   ii. Reflective optics.
   iii. Aspheric optical elements.
   iv. Liquid crystal spatial light modulators.
   v. Acousto-optical beam shaping.
   vi. Diffractive optics.
5. CBC and OPA.

Refractive and reflective optics result in relatively complex systems since they are based on spherical elements. Using aspherical components will reduce the number of elements, but these are more expensive to produce. Liquid crystal spatial light modulators are effective for real-time application, but limited to low powers (<250 W) since higher powers will damage the crystals. Acousto-optical beam shaping is still in the research stages. Diffractive optics, on the other hand, provide a great deal of flexibility in beam shaping, enabling a wide variety of intensity profiles to be produced economically. However, they are not effective for real-time application, since they are based on solid optical elements (hardware). In the next section, we discuss the basic concepts of beam shaping using diffractive optics, followed by the CBC and OPA technique.

3.8.1 Beam Shaping Using Diffractive Optics

Differences between refractive (classic) and diffractive optical elements (DOEs) are illustrated with the simple case in Figure 3.17, where it is evident that a diffractive lens can be generated from a refractive lens by removing rectangular sections of material that do not contribute to focusing of the light. The resulting diffractive lens is thus smaller in size, while maintaining the same focusing characteristics.

Factors that need to be considered in designing a DOE are the following:

1. Efficiency of transformation by the optical element.
2. Minimization of error sources such as chromatic aberration and thermal sensitivity.
3. Ease of fabrication.

Using diffractive optics for beam shaping involves designing a DOE that will generate a desired beam profile for a given input profile. A DOE essentially uses interference and phase modification to modulate a wave profile by first segmenting the wavefronts and redirecting the segments. We start our discussion on design of DOEs by considering the simple and well-known case of the Fresnel phase plate.

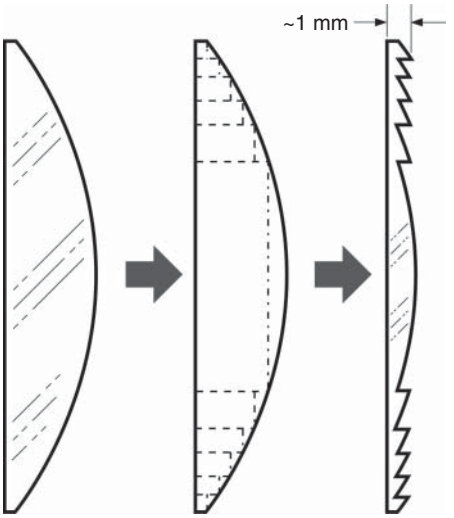


Figure 3.17 Illustration of how a diffractive optical element (to the right) can be obtained from a refractive element (to the left). The dashed rectangles are the parts removed.

3.8.1.1 Fresnel Phase Plate The Fresnel phase plate is a form of diffractive lens and, as such, diffracts and focuses a collimated monochromatic radiation at a point P. It is based on the concept of Fresnel zones where a circular aperture can be divided into annular regions such that each of the regions has the same area and contributes the same amount of light at *P*, which is located on the axis through the center of the aperture (Figure 3.18). The radii of the inner and outer rings of the annular regions are such that the distance from the periphery of each ring to the point *P* is equal to the distance from *P* to the aperture center (*f*) plus an integer number of half-wavelengths. The annular regions are the Fresnel zones and the radii of the zones can be shown (using Pythagoras' theorem) to be given by:

$$r_{\mathrm{m}}^2 = fm\lambda \tag{3.48}$$

where f is the focal length ($=r_{\mathrm{o}}$) and m is an integer.

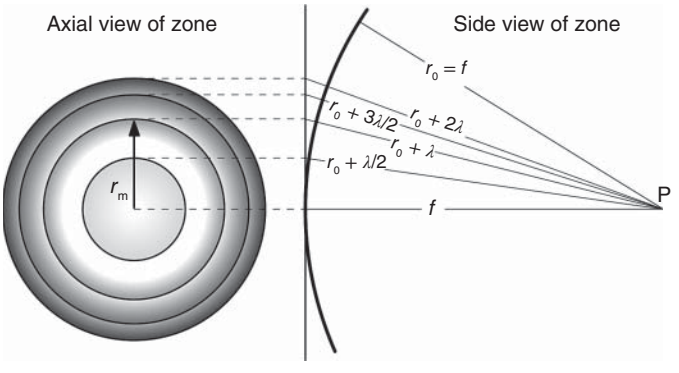


Figure 3.18 Fresnel zones.

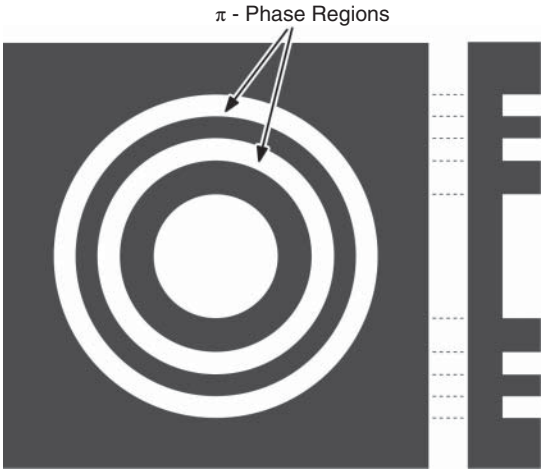


Figure 3.19 Fresnel phase plate.

Since the optical distance from contiguous Fresnel zones to P differs on average by a half-wavelength, the zones are 180° out of phase and thus block each other out in terms of the electric field contributed at P. The electric field at P will thus vary between two values as the aperture diameter is increased. However, by shifting the phase of alternating zones by 180°, the contiguous zones will now reinforce each other at P. This phase shift is achieved by using the DOE as shown in Figure 3.19, the Fresnel phase plate. The 180° (π) phase-shift regions are achieved with grooves of depth:

$$d = \frac{(2^N - 1)\lambda}{2^{N+1}(n_r - 1)} \tag{3.49}$$

where n_r is the refractive index of the DOE and N is the number of masks needed to manufacture the DOE. For the simple binary case shown in Figure 3.19, $N = 1$. As N increases, the element becomes more accurate. In the limit, N tends to infinity for a kinoform, and the depth equation then reduces to

$$d = \frac{\lambda}{2(n_r - 1)} \tag{3.50}$$

The simple binary (two-level) structure is not very efficient, and the efficiency can be improved by increasing the number of levels (multilevel) as shown in Figure 3.20 for two-, four-, and eight-level structures. The diffraction efficiency is the ratio of the power of the diffracted beam to the incident power and results from diffraction loss. Thus, a 60% phase plate will have only about 60% of the incident light being focused onto the axis. In the limit, the resulting continuous profile, the kinoform, has 100% efficiency. The phase functions for the different structures (phase levels) are shown in Table 3.1, together with the depth of etching necessary to achieve that phase function, and the corresponding efficiency.

3.8.1.2 Diffractive Optics Design For some manufacturing applications such as laser surface heat treatment, it is desirable to use a beam that has uniform irradiation distribution that drops off sharply at the boundary (flat top), since it provides an even distribution of the beam irradiation on the area being processed. Flat-top beams have also been used in illumination and skin treatments in dermatology. In general, customized beam shaping can be useful in a variety of applications, for example,

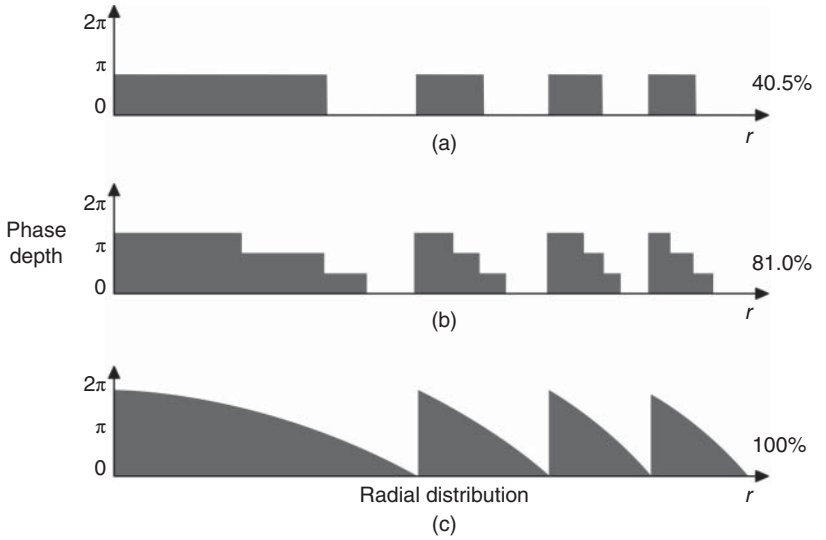


Figure 3.20 Multilevel Fresnel phase plate. (a) Two-level Fresnel phase plate. (b) Four-level. (c) Kinoform profile. Source: Adapted from Swanson and Veldkamp (1989).

Table 3.1 Phase functions for Fresnel phase plate.

No. of phase levels	Phase function $m = 0, \pm 1, \pm 2, \ldots$	Phase shift (etch depth)	Efficiency	Number of masks needed
2	$\varphi(x, y) = (m+1)\pi$	π	40.5	1
4	$\varphi(x, y) = (m+1)\frac{\pi}{2}$	$\pi/2$	81.0	2
8	$\varphi(x, y) = (m+1)\frac{\pi}{4}$	$\pi/4$	95.0	3
16	$\varphi(x, y) = (m+1)\frac{\pi}{8}$	$\pi/8$	99.0	4

Source: Adapted from Swanson and Veldkamp (1989).

laser welding of dissimilar materials. Even though in theory, diffractive optics can be used to transform any beam profile into any other profile, reality dictates that the input profile be as simple as possible. We shall therefore focus on the use of diffractive optics to transform Gaussian beams into flat-top beams (Figure 3.21), to illustrate the basic concepts. For such a case, it is essential for the input Gaussian (TEM_{00}) beam to be of high quality, with a low M^2 value of about 1.1 or less. Where the input beam is of lower quality, it may be necessary to use a spatial filter to enhance the beam quality before transformation.

One common technique for generating a flat-top beam from a Gaussian mode is to partition the original beam into sections which are then remapped unto the focal plane. The uniform profile is achieved by diverging the higher-intensity regions near the center and directing them away from the center in a controlled manner, while converging lower-intensity regions (Figure 3.22). To avoid spikiness due to interference effects, the sectioned beams are not superposed or made to overlap between the beam shaper and the focal plane. The drawback with this method is that it is only applicable to single-mode beams.

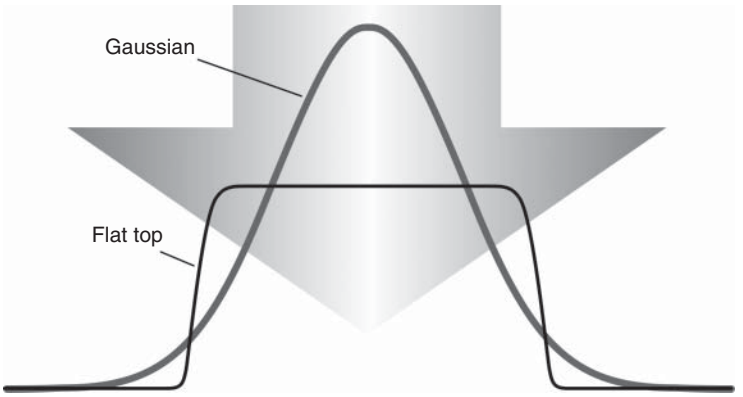


Figure 3.21 Basic beam shapes: Gaussian and Flat top.

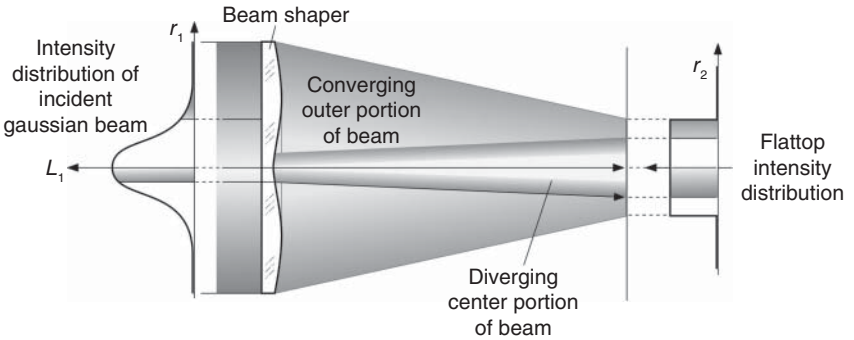


Figure 3.22 Generating a flat-top beam from a Gaussian beam by geometric mapping. Source: From Fuse (2015)/John Wiley & Sons.

DOEs have an advantage in the variety of beam shapes that they can generate. Generally, an optical element can transform an input wavefront into a desired output by modifying the wavefront phase profile, $\phi(r)$. For a phase-only element, this transformation, T, may be expressed as:

$$T(r) = e^{i\varphi(r)} \tag{3.51}$$

where r is the radial distance from the element center. $\phi(r)$ essentially defines the etch depth and zone transition points.

The phase profile of an optical element results from the phase delay (difference), $\Delta\phi$, at any level due to travel through the element. For a collimated beam traveling along the optical axis, the phase difference at any radial distance from the optical axis of the element is

$$\Delta\phi = \frac{2\pi n_r h}{\lambda} - \frac{2\pi h}{\lambda} = \frac{2\pi h}{\lambda}(n_r - 1) \tag{3.52}$$

where h is the thickness at any level and is a function of the radial distance, r.

Thus, one basic approach to generating a flat-top beam from a Gaussian beam is to use a phase grating as a multilevel DOE where transformation involves modulation on the phase of the input field. It can be shown that for a Gaussian beam normal to a transmissive DOE, the phase function that transforms it into a flat-top beam is

$$\varphi_x(x) = \frac{k_w d_x}{2z\ \operatorname{erf}\left(\frac{D_x}{\sqrt{2}w_x}\right)}\left[x\operatorname{erf}\left(\frac{\sqrt{2}x}{w_x}\right) + \frac{w_x\left(e^{-\frac{2x^2}{w_x{}^2}} - 1\right)}{\sqrt{2\pi}}\right] - \frac{k_w x^2}{2z} \tag{3.53}$$

$$\varphi_y(y) = \frac{k_w d_y}{2z\operatorname{erf}\left(\frac{D_y}{\sqrt{2}w_y}\right)}\left[y\operatorname{erf}\left(\frac{\sqrt{2}y}{w_y}\right) + \frac{w_y\left(e^{-\frac{2y^2}{w_y{}^2}} - 1\right)}{\sqrt{2\pi}}\right] - \frac{k_w y^2}{2z} \tag{3.54}$$

where $\phi_x(x)$ and $\varphi_y(y)$ are the phases in the x and y directions, respectively; d_x d_y dimensions of the rectangular flat-top; D_x D_y dimensions of the DOE; and w_x w_y beam waist dimensions. x and y directions define the plane of the DOE, and z is the direction of beam propagation.

A DOE with grating function that is the sum of $\phi_x(x)$ and $\varphi_y(y)$ will transform the Gaussian beam phase into a rectangular uniform profile (flat-top) beam.

For a circular beam, the corresponding phase function $\varphi(r)$ is given by:

$$\varphi(r) = \frac{k_w d}{2z\sqrt{1 - e^{-\frac{D^2}{2w^2}}}}\int_0^r \sqrt{1 - e^{-\frac{2\rho^2}{w^2}}}\,d\rho - \frac{k_w r^2}{2z} \tag{3.55}$$

where d and D are the flat-top distribution and DOE diameters, respectively, w is the beam waist size, and ρ is a dummy variable.

Line beam profiles have been shown to be more effective than the circular beams from which they are generated, in heat treating surface and thin films, since the surfaces can be more efficiently heated (up to about 20 times) than when a circular beam and 2D scanner are used.

There are commercial software packages by ZEMAX, 3333OSLO, and CODE V that facilitate the generation of the optical element geometry for a given phase function. The outcome of the transformation process depends on the beam wavelength, aperture (input beam size), working distance (or effective focal length), and output beam size.

3.8.1.3 Beam Propagation After Diffraction A given optical element design will generate the flat-top profile only at the focal plane. The beam profile changes with propagation distance (Figure 3.23), being convex or rounded at the top before the focal plane, and concave- or ring-shaped after the focal plane. This is due to errors (aberrations) associated with wavefront modulation.

To understand the effect of propagation on the diffracted beam, let us consider a DOE $E_l(x, y, 0)$ (Figure 3.24). The profile $E_l(X, Y, Z)$, at any general location, a distance Z from the aperture, can be related to the aperture profile, $E_l(x, y, 0)$, by the Rayleigh–Sommerfeld equation as follows:

$$E_l(X, Y, Z) = \frac{1}{2\pi}\iint_{x,y} E_l(x, y, 0)\frac{e^{-ik_w r}}{r}\frac{Z}{r}\left(ik_w + \frac{1}{r}\right)dxdy \tag{3.56}$$

where $r = [(x - X)^2 + (y - Y)^2 + Z^2]^{1/2}$. E_l represents the spatial component of the electric field.

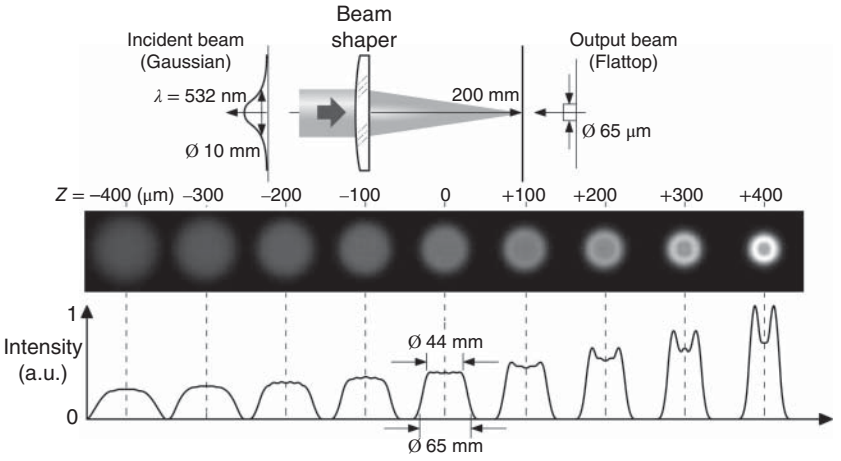


Figure 3.23 Variation of beam profile with propagation distance. Source: From Fuse (2015)/John Wiley & Sons.

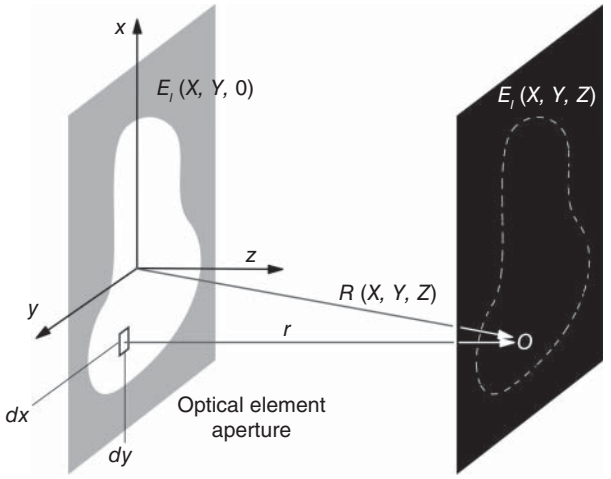


Figure 3.24 Propagation distance from aperture plane to the plane of interest.

Standard beam shapers that are based on DOEs are highly wavelength-dependent. Thus, a specific diffractive optics design will generate the desired profile over a narrow range of wavelengths. Polychromatic designs, on the other hand, are able to operate over a relatively broad wavelength range. However, the output profiles for such systems are not as accurate as those of the standard design.

3.8.1.4 Diffractive Optical Element Construction A number of techniques have been developed for constructing DOEs. Here, we outline the basic concepts of one common technique that is based on photolithography, using a simple design to illustrate the concepts. It involves the use of masks, photoresist, and etching, and the basic steps are illustrated in Figure 3.25a. The number of masks used depends on the number of phase levels desired (Table 3.1), and the relationship between them is

$$p = 2^N \tag{3.57}$$

where p is the number of phase levels and N is the number of masks.

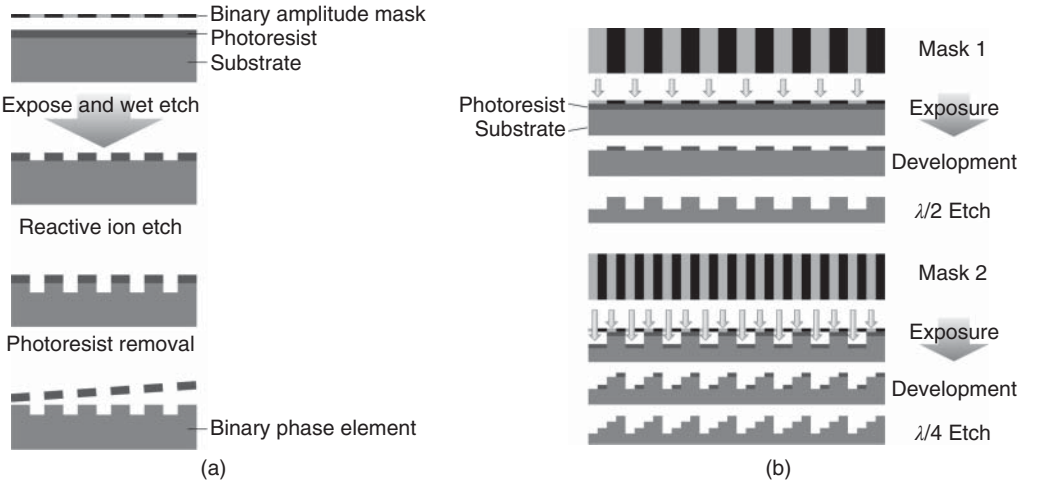


Figure 3.25 Production of a diffractive optical element profile: (a) binary masking and (b) multilevel masking. Source: Adapted from Swanson and Veldkamp (1989).

The mask can be produced using an electron beam pattern generator. It is first placed on a substrate that has a thin photoresist coating on the surface and then exposed to ultraviolet radiation. The photoresist is now developed, resulting in exposed and unexposed regions that are defined by the mask. Etching of the substrate now results in removal of the exposed regions in steps until the desired depth, which determines the optical path difference between the etched and unetched levels, is achieved. A finer profile can be achieved by using additional masks with smaller spacing and repeating the process in subsequent steps, as illustrated in Figure 3.25b.

3.8.2 Beam Shaping Using Coherent Beam Combining and Optical Phase Array

The CBC and OPA technique (or Civan method) is more suited for real-time beam shaping.

CBC enables high-power levels, up to 21 kW CW, to be obtained in single-mode fiber lasers. The principle involves splitting a seed laser into a number of channels (about 10 to over a 100), as shown in Figure 3.26. Each channel is then amplified individually, and the amplified beams are recombined into one high-power output. For real-time applications, recombination takes place in the far field using an interference pattern which is controlled electro-optically by rapid modulation of each channel's phase. This is the OPA stage which enables dynamic modulation of process parameters.

With OPA, all the channels are directed at the workpiece in the far field to create a specific interference pattern at the focal point by electro-optically manipulating the phases of the individual beams. The resulting interference pattern determines the beam properties such as 2D steering (wobbling), depth of field, shape, and power density.

The listed beam properties are currently manipulated using robots (1 Hz), mirrors (1 kHz), and DOEs (static). The response times of these elements are relatively slow, limiting the processing speeds. Furthermore, the use of mechanical components in some cases implies the need for maintenance. With electro-optical control in OPA, the parameters can be modulated at up to 10 MHz frequencies, making the process more suited for real-time control.

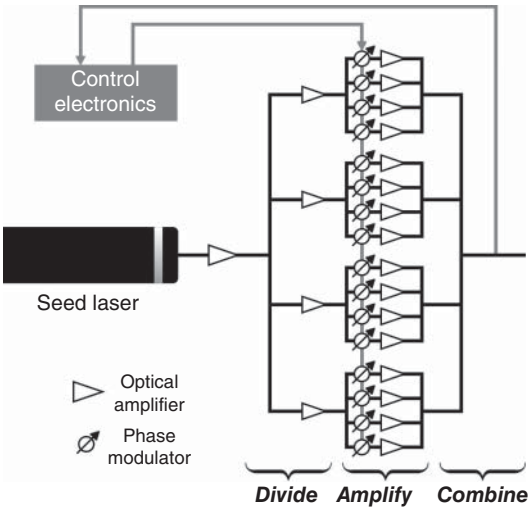


Figure 3.26 Principle of CBC technique. Source: From Shekel (2020)/Society of Photo-Optical Instrumentation Engineers.

3.9 SUMMARY

In electromagnetic wave theory, the product of the wavelength and frequency of electromagnetic radiation is equal to the velocity at which the radiation propagates in free space. When electromagnetic radiation or light beam enters a birefringent material, it is refracted into two rays, the ordinary and extraordinary rays. The ordinary ray obeys Snell's law and has the same velocity in any direction in the material, while the extraordinary ray has different velocities in different directions and does not follow Snell's law.

A light beam is said to be linearly or plane-polarized when the electric vector is restricted to a specific plane. When a light beam is incident on a surface, the electric vector associated with the beam can be resolved into two components, the p-component (lying in, or parallel to the plane of incidence) and the s-component (normal to the plane of incidence). The angle of incidence at which the reflected beam becomes plane-polarized normal to the plane of incidence is called the Brewster angle.

If two waves that have the same frequency and are plane-polarized in the same plane are superimposed on each other, the frequency of the resulting light vector is the same as that of the original light vectors. However, the amplitudes and phase angles are different. Now if the two plane waves lie in different planes which are orthogonal to each other, the resulting light vector will be elliptically polarized. Further, if the amplitudes of the two light vectors are equal, and the phase difference between them is 90°, the resulting light is circularly polarized.

Beam expanders are usually used to increase the size of a beam before focusing, and that reduces the size of the focused beam, increasing the power density at the focal point. This can be achieved using transmissive or reflective optics. On the other hand, beam splitters are used to split a light beam into two or more beams and can be done using either a wedge-shaped reflector or a partially reflective and partially transmissive optical element.

In this chapter, we have discussed the two major ways in which a beam can be delivered from the laser generator to the workpiece:

1. Using a conventional combination of transmissive and reflective optics (lenses and mirrors)
2. Using a fiber optic cable (which will usually also include some lenses).

Major components of a conventional beam delivery system include the beam bending assembly, focusing assembly, interconnecting beam guard tubes, and different types of mirrors and lenses.

Beam delivery can also be achieved using fiber optic systems, even though these may not be applicable to all types of high-power lasers. The basic structure consists of the transmitter, transmission line (of interest in processing), and receiver. The waveguide, in which the light beam propagates, consists of two principal components: core and cladding. The core is the inner portion through which light is transmitted, and the cladding is the concentric outer layer with a relatively lower index of refraction. The waveguide is characterized by the numerical aperture, normalized frequency, and angle of acceptance. The major fiber types are the step-index multimode, single-mode, and graded-index multimode fibers. Fiber optic systems are normally limited by beam degradation, and the major forms of degradation are dispersion or distortion (including chromatic or material dispersion, waveguide dispersion, and modal dispersion), and optical losses such as input-coupling losses, connector/splice losses, fiber losses, and output-coupling losses.

The profile or shape of a laser beam has significant influence on the process outcome. There are different techniques that can be used for changing the beam profile, but one of the most advantageous involves the use of diffractive optics. These change the profile by modulating the phase as a result of phase delay at any level due to travel through the DOE. The element design can thus be generated once the phase profile is known. Special software packages are available for this step. The designed optical element can then be produced by a process that is based on photolithography and involves the use of masks, photoresist, and etching. A DOE is most effective for the specific wavelength for which it was designed. Changing the beam wavelength will deteriorate the profile. Furthermore, the desired profile is obtained at the focal point and is modified outside this region (before or after).

The basic concepts of beam shaping using the CBC and OPA technique are then discussed. With capability for MHz frequency modulation, this method is highly suited to real-time applications.

APPENDIX 3.A

List of symbols used in the chapter

Symbol	Parameter	Units
b_c	Normalized propagation constant	—
d	Fresnel plate groove depth	μm
d_x, d_y	Dimensions of the rectangular flat top	m
D_d	Dispersion coefficient	ps/(km nm)
D_w	Dimensionless dispersion coefficient	—
D_x, D_y	Dimensions of the DOE	m
E_l	General amplitude of electric field	V/m
$f = r_0$	Focal length	m
h	Thickness	m
k_1	Constant	—
$k_{w0} = \frac{2\pi}{\lambda_0}$	Wave number in free space	/m

Symbol	Parameter	Units
K_{ij}	Constant	dB/km
K_{Rs}	Rayleigh scattering coefficient	(dB/km) μm^4
L_{f}	Fiber length	m
m	Integer	—
n_1	Index of refraction of core	—
n_2	Index of refraction of cladding	—
n_{g}	Group index of refraction	—
n_{i}	Index of refraction of medium i	—
n_{r}	Core index of refraction at radius r	—
n_{ij}	Index of refraction of medium i relative to medium j	—
N	Number of masks needed to manufacture a DOE	—
p	Number of phase levels	—
q_{in}	Input power	W (Watts)
q_{out}	Output power	W (Watts)
r_{c}	Core radius	m
r_{m}	Fresnel zone radius	m
R_{p}	Reflection coefficient of ray in plane of incidence	—
R_{s}	Reflection coefficient of ray normal to plane of incidence	—
t_{max}	Travel time of the slowest mode	s (seconds)
t_{min}	Travel time of the fastest mode	s (seconds)
T	Transformation	—
u_{g}	Group velocity	m/s
u_{w}	Wave velocity	m/s
w_x, w_y	Beam waist dimensions	m
x, y	Define plane of the DOE	m
z	Direction of beam propagation	m
α_{l}	Attenuation	dB/km
α_{ir}	Attenuation due to infrared absorption	dB/km
α_{Rs}	Attenuation due to Rayleigh scattering	dB/km
α_{uv}	Attenuation due to ultraviolet absorption	dB/km
γ_{c}	Rate of decay of electric field in cladding	/m
λ_{o}	Wavelength in free space	m
χ_{g}	Group delay	s/m
$\varepsilon_{\mathrm{p0}}$	Electric permittivity in free space	farad/m (F/m)
$\phi(x,y)$	Phase function	—
μ_{m0}	Magnetic permeability in free space	henry/m (H/m)
ν_{n}	Normalized frequency	—
Δ	$\frac{n_1 - n_2}{n_1}$	—
$\Delta\phi$	Phase delay (difference)	—
$\Delta\tau_{\mathrm{m}}$	Difference in arrival times of mode with largest waveguide group delay and that with least delay	s
$\Delta\tau_{\mathrm{cd}}$	Pulse broadening or delay distortion	s
$\Delta\tau_{\mathrm{wd}}$	Pulse broadening due to waveguide dispersion	s
θ_1	Angle of incidence	°
θ_2	Angle of refraction	°
θ_{a}	Angle of acceptance	°
θ_{c}	Critical angle of incidence	°

PROBLEMS

3.1. Given two plane-polarized waves which are polarized in orthogonal planes and are superimposed on each other,

$$E_{ly}(t) = E_{0y}\cos(\varphi_y - \omega t)$$

and

$$E_{lz}(t) = E_{0z}\cos(\varphi_z - \omega t)$$

show that

$$\left[\frac{E_{ly}(t)}{E_{0y}}\right]^2 - 2\frac{E_{ly}(t)E_{lz}(t)}{E_{0y}E_{0z}}\cos(\varphi_z - \varphi_y) + \left[\frac{E_{lz}(t)}{E_{0z}}\right]^2 = \sin^2(\varphi_z - \varphi_y)$$

3.2. Given two plane-polarized waves which are polarized in the same plane and are superimposed on each other,

$$E_{l1}(t) = E_{01}\cos(\varphi_1 - \omega t)$$

and

$$E_{l2}(t) = E_{02}\cos(\varphi_2 - \omega t)$$

show that

$$E_l(t) = E_0\cos(\varphi - \omega t)$$

where

$${E_0}^2 = {E_{01}}^2 + {E_{02}}^2 + 2E_{01}E_{02}\cos(\varphi_2 - \varphi_1)$$

and

$$\tan\varphi = \frac{E_{01}\sin\varphi_1 + E_{02}\sin\varphi_2}{E_{01}\cos\varphi_1 + E_{02}\cos\varphi_2}$$

3.3. Given that

$$NA = \sin\theta_a$$

and assuming that the refractive index for air is 1, show that for step-index fibers, NA may also be expressed as:

$$NA = \sqrt{{n_1}^2 - {n_2}^2}$$

3.4. Determine the core radius and refractive index of a step-index multimode fiber that has a normalized frequency of 75 and a numerical aperture of 0.25. Assume the wavelength of the propagating medium is 1.06 μm and that the cladding refractive index is 1.52.

3.5. Using Eq. (3.33), obtain an expression for the group index, n_g, in terms of the wavelength λ.

3.6. Show that the normalized frequency, ν_n, can also be expressed as:

$$\nu_n = k_w r_c n_1\sqrt{2\Delta - \Delta^2}$$

3.7. (a) The refractive index of glass is 1.52 and that of water is 1.33. If light travels from glass into water, determine the resulting Brewster and critical angles.
(b) Consider the He-Ne laser shown in Figure 2.3. What should the angle θ_B be if the light reflected from the windows is to be totally polarized perpendicular to the plane of incidence? Assume the windows are made of glass of refractive index 1.52.

3.8. The core of a step-index fiber optic cable which is in air has a refractive index of 1.62, while the cladding has an index of 1.52. What is the entrance cone angle for light that enters the fiber? What will be the corresponding numerical aperture? Determine the modal dispersion time delay per kilometer, for the fiber.

3.9. (a) Given

$$\nu\lambda = c$$

show that for small $\Delta\lambda$,

$$\Delta\nu = -\left(\frac{c}{\lambda^2}\right)\Delta\lambda$$

(b) Given a frequency bandwidth of $\Delta\nu = 5500$ Hz, determine the corresponding wavelength bandwidth at a wavelength of $\lambda = 0.8$ μm.

3.10. A multimode waveguide has a core refractive index of 1.5, a core radius of 30 μm, and a Δ of 1.5%. Determine the numerical aperture and normalized frequency for a laser beam of wavelength $\lambda = 1.06$ μm that enters the waveguide.

3.11. For the setup in Problem (3.10), determine how the number of modes in the waveguide will vary as

(a) The core refractive index
  (i) increases
  (ii) decreases
(b) The incident wavelength
  (i) increases
  (ii) decreases.

3.12. A single-mode step-index waveguide has a core refractive index of 1.52, a cladding refractive index of 1.50, a normalized frequency of 2.0, and an input signal linewidth of $\Delta\lambda = 0.002$ μm. If the input signal wavelength is $\lambda = 1.06$ μm, determine the delay distortion per kilometer of fiber waveguide.

3.13. (a) An Nd:YAG laser beam is delivered to a workpiece through an optical fiber cable of length 5 m. If the power that is measured at the workpiece is 500 W, determine the power output of the laser, assuming the only losses are those incurred in the fiber cable.
Hint: Use Figure 3.16.
(b) What would be the loss in dB/km if 95% of the beam was delivered?

3.14. An Nd:YAG laser beam is delivered to a workpiece using conventional beam delivery as shown in Figure P3.14.
(a) Identify the components (for example, beam splitter, beam expander, lens) labeled A, B, and C in the path of the laser beam and sketch the schematic of the components. Explain why component A is used in the delivery system.
(b) Assuming that the total focal length of component A is 210 mm, calculate the focal lengths or focal length ratio of all the lenses used in components A, B, and C.

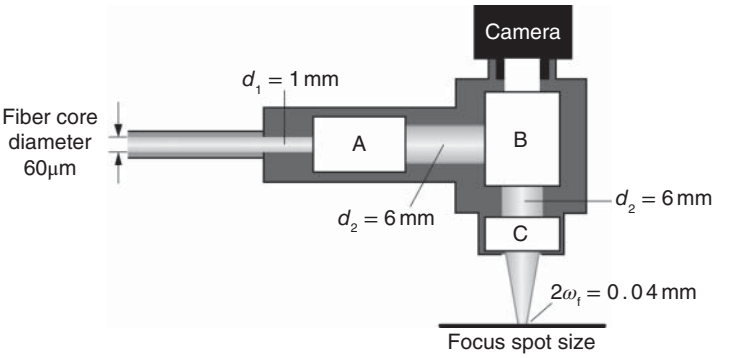


Figure P3.14

3.15. (a) If the Nd:YAG laser beam in Problem (3.14) is delivered through a step-index multimode optical fiber, determine the approximate total number of modes M_c. The refractive indices of the core and cladding for the step-index multimode fiber are $n_1 = 1.6$, $n_2 = 1.5$.

(b) Determine the range of the optical fiber cable length so that the power measured at the workpiece is no less than 1000 W and the pulse broadening time is no more 300 ns. The power input at the beginning of the fiber cable is 1200 W, and Figure 3.16 may be used for this problem.

BIBLIOGRAPHY

Buck, J.A. (1995). *Fundamentals of Optical Fibers*. New York: Wiley.

Cherin, A.H. (1983). *An Introduction to Optical Fibers*. New York: McGraw-Hill.

Chryssolouris, G. (1991). *Laser Machining*. Theory and Practice, Berlin: Springer-Verlag.

Dally, J.W. and Riley, W.F. (1991). *Experimental Stress Analysis*, Thirde. New York: McGraw Hill.

Fuse, K. (2015). Beam shaping for advanced laser materials processing. *Laser Technik Journal* 2015: 19–22.

Gnanamuthu, D.S. and Shankar, V.S. (1985). Laser heat treatement of iron-base alloys. *SPIE, Applications of High Powered Lasers* 572: 56–72.

Keiser, G. (1991). *Optical Fiber Communications*. New York, NY: McGraw-Hill.

Killen, H.B. (1991). *Fiber Optic Communications*. Englewood Cliffs, NJ: Prentice Hall.

Kleekamp, C. and Metcalf, B. (1978). *Designer's Guide to Fiber Optics - Part 1*. Boston, MA: Cahners Publishing Company.

Kosoburd, T. and Kedmi, J. (1992). Beam shaping with diffractive optical elements. *SPIE, 8th Meeting on Optical Engineering in Israel* 1971: 390–399.

Lachs, G. (1998). *Fiber Optic Communications*. New York, NY: McGraw-Hill Telecommunications.

Lacy, E.A. (1982). *Fiber Optics*. Englewood Cliffs, NJ: Prentice Hall.

Leong, K.H. and Hunter, B.V. (1996). High-power fiberoptic laser beam delivery moves forward. *Industrial Laser Review* 6: 7–12.

Luxon, J.T. and Parker, D.E. (1985). *Industial Lasers and Their Applications*. Englewood Cliffs, New Jersey: Prentice Hall.

Račiukaitis, G., Stankevičius, E., Gečys, P. et al. (2011). Laser processing by using diffractive optical laser beam shaping technique. *JLMN-Journal of Laser Micro/Nanoengineering* 6 (1): 37–43.

Shekel, E., Vidne, Y., and Urbach, B. (2020). 16 kW single mode CW laser with dynamic beam for material processing, *Proc. SPIE 11260, Fiber Lasers XVII: Technology and Systems*, 1126021 (21 February 2020); https://doi.org/10.1117/12.2545900.

Swanson, G.J. and Veldkamp, W.B. (1989). Diffractive optical elements for use in infrared systems. *Optical Engineering* 28 (6): 605–608.

Ungar, S. (1990). *Fiber Optics, Theory and Applications*. New York: Wiley.

Wilson, J. and Hawkes, J.F.B. (1987). *Lasers - Principles and Applications*. New York: Prentice Hall International Series in Optoelectronics.

PART II
Engineering Background

4 Heat and Fluid Flow

The heat and fluid flows that occur during laser processing influence the microstructures (through the grain structure and phases that are formed), residual stresses (through the thermal stresses that result from differential strains), and distortion that evolves during the process. These in turn affect the mechanical properties and thus quality of the process. The discussion in this chapter is in three sections. In the first section, simplified lumped parameter energy balance equations are presented that enable quick estimates to be made of energy requirements for a given process. Then in the second section, only heat flow in the solid part of the workpiece is considered. The analysis in this section is applicable, as a first approximation, to all the processes discussed, and is more accurate in the solid part of the workpiece. To enable closed-form solutions to be obtained, the heat source is initially modeled as either a point or line source. Analytical expressions are then obtained for the temperature distribution for the quasi-stationary state, together with the cooling rates, and peak temperatures. A Gaussian distribution, which is more representative of the heat source, is then considered. This is followed by the two-temperature model, which is appropriate for ultrashort pulse laser beams in metals. In the third section, flow in the molten pool for processes that involve melting (such as welding, cladding, surface melting, and cutting) is discussed. In the second and third sections, the basic governing equations are first presented. Then where feasible, approximate closed-form analytical solutions are presented that enable a quick analysis of the process to be made.

4.1 ENERGY BALANCE DURING PROCESSING

During processing, part of the beam that is incident on the workpiece is reflected away, and the remaining part is absorbed by the workpiece. The absorbed energy, Q_a, is what is used in processing and may involve

1. Heating up the solid to the melting temperature, $Q_1 = m_a c_p \Delta T_m$
2. Melting, which requires the latent heat of melting, $Q_2 = m_a L_m$
3. Heating up the molten material to the vaporization temperature, $Q_3 = m_a c_p \Delta T_v$
4. Vaporization, which requires the latent heat of vaporization, $Q_4 = m_a L_v$
5. Energy lost by conduction, convection, and radiation, Q_l.

where c_p is the specific heat (J/kg K), L_m is the latent heat of melting per unit mass (J/kg), L_v is the latent heat of vaporization per unit mass (J/kg), m_a is the mass (kg), T_m is the melting temperature (°C), T_v is the vaporization temperature (°C), $\Delta T_m = T_m - T_0$, and $\Delta T_v = T_v - T_m$

Thus, an energy balance gives

$$Q_a = m_a c_p \Delta T_m + m_a L_m + m_a c_p \Delta T_v + m_a L_v + Q_l \tag{4.1a}$$

Principles of Laser Materials Processing: Developments and Applications, Second Edition. Elijah Kannatey-Asibu, Jr.
© 2023 John Wiley & Sons, Inc. Published 2023 by John Wiley & Sons, Inc.
Companion Website: www.wiley.com/go/Kannatey-Asibu/PrinicplesLaserMaterialsProcessing

Assuming all the absorbed energy is used in processing, and neglecting any losses, we have the following energy balance for the process:

$$Q_a = m_a c_p \Delta T_m + m_a L_m + m_a c_p \Delta T_v + m_a L_v = \rho V(c_p \Delta T_m + L_m + c_p \Delta T_v + L_v) \tag{4.1b}$$

where ρ is density (kg/m^3) and V is volume (m^3).

Since part of the absorbed energy is conducted into the surrounding material, it experiences a temperature rise, with subsequent cooling. Conduction into the workpiece is discussed in the next section.

4.2 HEAT FLOW IN THE WORKPIECE

The principal thermal factors that determine the microstructure of the workpiece are the temperature distribution, the peak temperatures, and the cooling rates that are experienced during the process. Once an expression for the temperature distribution is obtained, it can be used to derive the peak temperature and cooling rate expressions. The following analysis of the heat flow in the solid workpiece is therefore discussed along these lines.

4.2.1 Temperature Distribution

The discussion on the temperature distribution starts with the governing equation for conduction heat flow. This is based on the assumption that heat conduction through the workpiece is usually much greater (for materials with very low thermal conductivity, this assumption can lead to significant error and thus might need to be reconsidered) than any heat exchange with the surroundings by natural convection or radiation. It is further assumed that the workpiece surfaces are adiabatic, that is, there is no heat loss or gain by either convection or radiation.

With these assumptions, the governing equation for heat flow in a solid with a coordinate system fixed at a stationary origin relative to the solid is given (Appendix 4.B) by:

$$\rho c_p \frac{dT}{dt} = \frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial y}\left(k\frac{\partial T}{\partial y}\right) + \frac{\partial}{\partial z}\left(k\frac{\partial T}{\partial z}\right) + q_s \tag{4.2a}$$

$$\Rightarrow \rho c_p \frac{dT}{dt} = k\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}\right) + \frac{\partial k}{\partial T}\left[\left(\frac{\partial T}{\partial x}\right)^2 + \left(\frac{\partial T}{\partial y}\right)^2 + \left(\frac{\partial T}{\partial z}\right)^2\right] + q_s \tag{4.2b}$$

or in a more compact form:

$$\rho c_p \frac{dT}{dt} = \nabla \cdot (k \nabla T) + q_s \tag{4.2c}$$

where k is the thermal conductivity (W/mm K), q_s is the rate of local internal energy generated per unit volume (W/mm^3), t is the time (s), and T is the temperature (°C).

Equation (4.2) is nonlinear, and thus it is difficult to solve in closed form. To make the problem more tractable analytically, the following simplifications are made:

1. The equation is linearized by assuming that the material's physical coefficients such as thermal conductivity are independent of temperature. Thus, $\partial k/\partial T = 0$.
2. The internal heat generation is neglected. This means $q_s = 0$. This assumption is reasonable for a number of applications, especially when one compares the external heat sources associated with some laser processes, with any heat that might be generated within the material. However,

this is not necessarily true of oxygen-assisted laser cutting where the exothermic reaction can be considerable.

3. The workpiece material is assumed to be homogeneous and isotropic. This is a reasonable approximation for many non-composite materials which have not been work-hardened.

With these simplifications, Eq. (4.2a, b) reduces to the form:

$$\frac{dT}{dt} = \kappa \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} \right) = \kappa \nabla^2 T \tag{4.3}$$

where $\kappa = k/\rho c_p$ is the thermal diffusivity (mm^2/s).

Solution of Eq. (4.3) requires six boundary conditions, two in each direction, and an initial condition.

One general approach to solving Eq. (4.3) is to use separation of variables by first expressing $T(x, y, z, t)$ as:

$$T = X(x)Y(y)Z(z)\tau(t) \tag{4.4}$$

However, before further discussing the solution of Eq. (4.3), we first consider possible forms of the heat flow problem in laser processing to see if the analysis can be further simplified. The heat source used in laser processing is usually one of two possible configurations:

1. A stationary source, for example, as used in laser drilling or spot welding, and such processes may be directly analyzed using the preceding equations.
2. A moving source, for example, as used in laser cutting, welding, or heat treatment.

Analysis of the moving heat source case is facilitated by using a coordinate system that is attached to the heat source. Let us therefore consider a coordinate system moving with the heat source along the x-axis, as shown in Figure 4.1. The corresponding governing equation is obtained by a coordinate transformation from the plate to the heat source (Appendix 4.C), with x being replaced by ξ, y by y', z by z', and t by t', that is,

$$\xi = x - u_x t, \quad y' = y, \quad z' = z, \quad t' = t$$

where u_x is the traverse velocity of the heat source in the x-direction (mm/s).

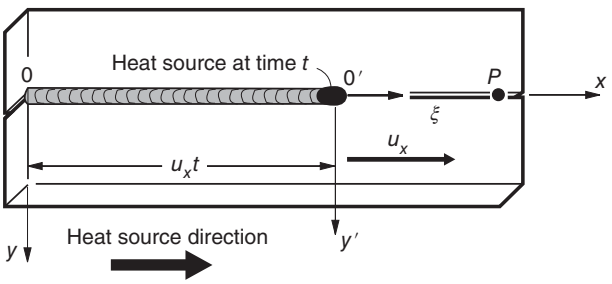


Figure 4.1 Schematic of moving coordinate system associated with laser processing.

The transformation gives the governing Eq. (4.3) in terms of the moving coordinate system (Appendix 4.C) as:

$$\frac{\partial T}{\partial t} = \kappa \left(\frac{\partial^2 T}{\partial \xi^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} \right) + u_x \frac{\partial T}{\partial \xi} \tag{4.5}$$

where ξ, y, z is a coordinate system attached to the moving heat source, with positive ξ in the direction in which the heat source is moving; x, y, z is a coordinate system with origin O and fixed to the workpiece, with positive x in the direction in which the heat source is moving; and $\partial T/\partial t$ is the time rate of change of temperature in the moving coordinate system.

When a uniform prismatic workpiece is processed over a long enough period, experimental results indicate that a state is reached when an observer positioned at the heat source or moving origin detects no change in the temperature distribution around the source. This is the *quasi-stationary state*, and for this case, the temperature undergoes no change with time with respect to the coordinate system attached to the heat source. The time derivative in Eq. (4.5) thus vanishes and the governing equation then becomes

$$\frac{\partial^2 T}{\partial \xi^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} = -\frac{u_x}{\kappa} \frac{\partial T}{\partial \xi} \tag{4.6}$$

This is the differential equation for the quasi-stationary case.

A convenient way of solving an equation of this form is to assume that T is of the form:

$$T = T_0 + \exp\left(-\frac{u_x \xi}{2\kappa} \right) \varphi(\xi, y, z) \tag{4.7}$$

where T_0 is the initial temperature of the workpiece and $\phi(\xi, y, z)$ is a function that is yet to be determined.

Substituting Eq. (4.7) into (4.6) results in the following equation:

$$\frac{\partial^2 \varphi}{\partial \xi^2} + \frac{\partial^2 \varphi}{\partial y^2} + \frac{\partial^2 \varphi}{\partial z^2} - \left(\frac{u_x}{2\kappa} \right)^2 \varphi = 0 \tag{4.8}$$

Similar differential equations are encountered in the analysis of electric wave problems. Thus, in this form, appropriate solutions are more easily obtained, depending on the boundary conditions that are imposed. To facilitate the analysis and obtain a closed-form solution of the preceding equation, the following additional simplifying assumptions are made in this section:

1. A point or line heat source is used. This assumption enables a closed-form solution to be obtained. A more accurate representation would be that of a Gaussian distribution. However, for that, we would have to resort to numerical methods.
2. No phase changes occur, that is, the effect of latent heat of fusion is negligible. As a first-order approximation, this is a reasonable assumption, since latent heat is absorbed in front of the heat source and evolved behind it.

The Gaussian heat source is considered separately in Section 4.2.4. Thus, for now, we consider two forms of solutions:

1. One case is that of a thick plate on which a point heat source moves and involves three-dimensional heat flow. This might be the case, for example, in conduction mode welding (Section 10.3.1).

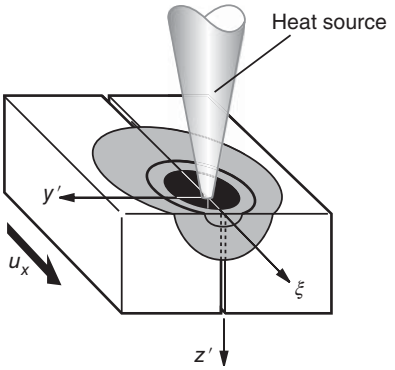


Figure 4.2 Three-dimensional configuration.

2. The other case is that of a thin plate with a line heat source that penetrates through the thickness and involves two-dimensional heat flow. Examples would be keyhole welding (Section 10.3.2) or laser cutting (Chapter 9).

To determine whether a plate is thin or thick, the following equation may be used as an initial approximation:

$$\beta_c = h\sqrt{\frac{\rho c_p u_x (T - T_0)}{q}} \tag{4.9}$$

where h is the plate thickness (m) and q is the heat flux (power) input (W).

The plate is considered to be thin when $\beta_c < 0.6$ and thick when $\beta_c > 0.9$. When high accuracy is desired and $0.6 < \beta_c < 0.9$, then it is best to solve the equations numerically. We now consider the closed-form solutions for the thick and thin plate cases separately.

4.2.1.1 Thick Plate with Point Heat Source (Three-dimensional) Figure 4.2 schematically illustrates this configuration. Ideally, the plate is considered to be semi-infinite, and the heat source to be a point. Now let q = heat flux (power) input and $r = \sqrt{\xi^2 + y^2 + z^2}$, where r is the distance of any general point, $P(\xi, y, z)$, from the moving origin (Figure 4.1).

However, to make the analysis even simpler, we start by considering the fully infinite case. Since the temperature of the plate remains unchanged far away from the source, the boundary conditions for this case can be constructed as follows:

$$\frac{\partial T}{\partial \xi} \to 0 \quad \text{as} \quad \xi \to \pm\infty, \quad \frac{\partial T}{\partial y} \to 0 \quad \text{as} \quad y \to \pm\infty, \quad \frac{\partial T}{\partial z} \to 0 \quad \text{as} \quad z \to \pm\infty \tag{4.10}$$

To account for the heat input, we consider the fact that for the point source, the heat flux through the surface of a sphere of radius r drawn around the source approaches q as $r \to 0$, that is,

$$\lim_{r \to 0} -4\pi r^2 \times k \times \frac{\partial T}{\partial r} = q \tag{4.11}$$

Now we have to obtain the solution for Eq. (4.8) which satisfies the preceding boundary conditions and accounts for the heat input as given in Eq. (4.11). Due to the nature of the boundary conditions

and the heat input equation, as well as the symmetrical nature of Eq. (4.8), ϕ depends only on r. Thus, using polar coordinates, Eq. (4.8) can be re-expressed as:

$$\frac{d^2\varphi}{dr^2} + \frac{2}{r}\frac{d\varphi}{dr} - \left(\frac{u_x}{2\kappa}\right)^2 \varphi = 0 \tag{4.12}$$

But

$$\frac{1}{r}\frac{d^2(\varphi r)}{dr^2} = \frac{d^2\varphi}{dr^2} + \frac{2}{r}\frac{d\varphi}{dr} \tag{4.13}$$

Thus, Eq. (4.12) becomes

$$\frac{d^2(\phi r)}{dr^2} = \left(\frac{u_x}{2\kappa}\right)\phi r = 0 \tag{4.14}$$

An obvious solution of Eq. (4.14) which also satisfies the boundary conditions is

$$\varphi r = k_1 e^{-\left(\frac{u_x r}{2\kappa}\right)} \quad \text{or} \quad \varphi = k_1 \frac{e^{-\left(\frac{u_x r}{2\kappa}\right)}}{r} \tag{4.15a}$$

where k_1 is to be determined.

This solution, Eq. (4.15a), also satisfies the heat input condition, Eq. (4.11), since

$$\frac{d\varphi}{dr} \times r^2 \rightarrow \quad \text{constant value as} \quad r \rightarrow 0 \tag{4.15b}$$

Now substituting Eq. (4.15a) into (4.7) gives the expression for the temperature distribution for the infinite case as:

$$T - T_0 = \frac{q}{4\pi k r} \exp\left(-\frac{u_x(r+\xi)}{2\kappa}\right) \tag{4.16}$$

The corresponding solution for the semi-infinite case, which is more representative of laser processing of an infinite plate (semi-infinite space), follows from Eq. (4.16) as:

$$T - T_0 = \frac{q}{4\pi k r} \exp\left[-\frac{u_x(r+\xi)}{2\kappa}\right] \tag{4.17}$$

The factor 2 that results in Eq. (4.17) is due to the fact that we are now considering only one half of the space. For laser processes that involve melting, Eq. (4.17) applies only outside the fused zone, that is, for $T < T_m$ (the melting temperature), and is more accurate farther away from the fusion boundary.

4.2.1.2 Thin Plate with Line Heat Source (Two-dimensional) The configuration for this case is illustrated schematically in Figure 4.3. In this case, heat flow is in two directions, ξ (or x)-direction, and y-direction. There is no flow in the z-direction. The heat source is considered to be a line that goes through the entire plate thickness uniformly. Thus, heat is input to the system as power per unit thickness. Now let $r = \sqrt{\xi^2 + y^2}$, the radius of a cylinder drawn around the heat source.

Since the heat source is uniform through the thickness, there can be no change in temperature in the thickness direction. Thus, we have

$$\frac{\partial T}{\partial z} = 0 \quad \text{for all } z$$

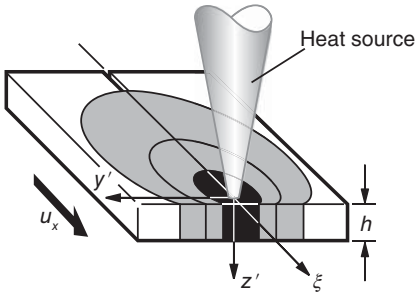


Figure 4.3 Two-dimensional configuration.

Eq. (4.8) then becomes

$$\frac{\partial^2\varphi}{\partial\xi^2} + \frac{\partial^2\varphi}{\partial y^2} - \left(\frac{u_x}{2\kappa}\right)^2\varphi = 0 \tag{4.18}$$

Again since the temperature of the plate remains unchanged far away from the source, the boundary conditions for this case can be constructed as follows:

$$\frac{\partial T}{\partial\xi} \to 0 \quad \text{as} \quad \xi \to \pm\infty, \quad \frac{\partial T}{\partial y} \to 0 \quad \text{as} \quad y \pm \infty \tag{4.19}$$

And yet again to account for the heat input, we consider the fact that since this is a line source, the heat flux through the surface of a cylinder drawn around the line source approaches q as $r \to 0$, that is,

$$\lim_{r\to 0} -2\pi rhk \times \frac{\partial T}{\partial r} = q \tag{4.20}$$

Since Eq. (4.18) is symmetrical in nature, and ϕ depends only on r, it can be expressed in cylindrical coordinates as:

$$\frac{d^2\varphi}{dr^2} + \frac{1}{r}\frac{d\varphi}{dr} - \left(\frac{u_x}{2\kappa}\right)^2\varphi = 0 \tag{4.21}$$

The solution to Eq. (4.21) that satisfies the boundary conditions is the modified Bessel function of the second kind and zero order, $K_0(u_x r/2\kappa)$.

It can be shown that

$$K_0\left(\frac{u_x r}{2\kappa}\right) \to \ln r \quad \text{as} \quad r \to 0$$

Thus,

$$\frac{dK_0}{dr} \to \quad \text{constant value as} \quad r \to 0$$

Eq. (4.20) is therefore satisfied by $K_0(u_x r/2\kappa)$.

Furthermore,

$$K_0\left(\frac{u_x r}{2\kappa}\right) \to \sqrt{\frac{\pi}{\left(\frac{u_x r}{\kappa}\right)}}\, e^{-\left(\frac{u_x r}{2\kappa}\right)} \quad \text{as} \quad r \to \infty$$

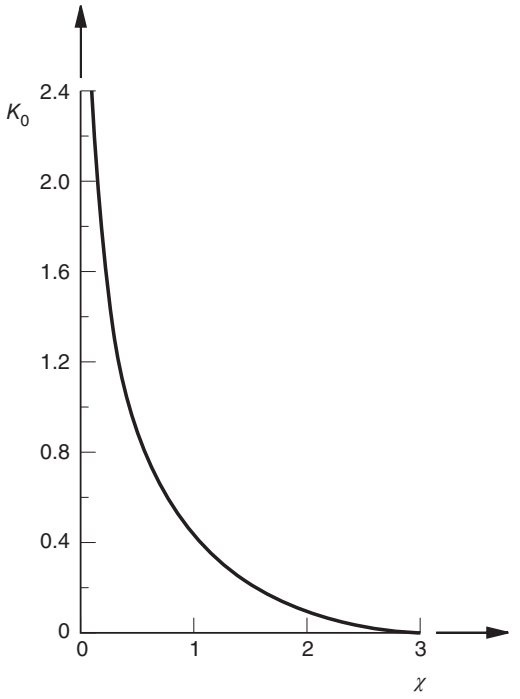


Figure 4.4 Plot of the modified Bessel function of the second kind of order zero, $K_0(\chi)$ (see also Appendix 4.G). Source: Adapted from Abramowitz and Stegun (1964).

Thus, the boundary conditions, Eq. (4.19) are satisfied. The solution for the thin plate case can therefore be expressed as:

$$T - T_0 = \frac{q}{2\pi kh} \exp\left(-\frac{u_x \xi}{2\kappa}\right) \times K_0\left(\frac{u_x r}{2\kappa}\right) \tag{4.22}$$

where $K_0(\chi)$ is the modified Bessel function of the second kind of order zero, as shown in Figure 4.4 and Appendix 4.G, and χ is the argument of the modified Bessel function.

When χ is small,

$$K_0(\chi) \approx -\left(0.5772 + \log\frac{\chi}{2}\right)$$

When χ is large,

$$K_0(\chi) \approx \sqrt{\left(\frac{\pi}{2\chi}\right)} e^{-\chi}$$

Equations (4.17) and (4.22) are also sometimes known as the Rosenthal equations, after the person who first derived them. Samples of the temperature distribution as represented by a family of isotherms drawn around the instantaneous heat source position (x–y plane) are shown in Figures 4.5–4.7. Figure 4.5 shows the effect of thermal conductivity by comparing the isotherms for a relatively low thermal conductivity material (say steel), and a relatively high thermal conductivity material (say aluminum) when other processing conditions are the same. Figure 4.6 shows the effect of speed on the isotherms, for the same input power. Finally, Figure 4.7 compares the

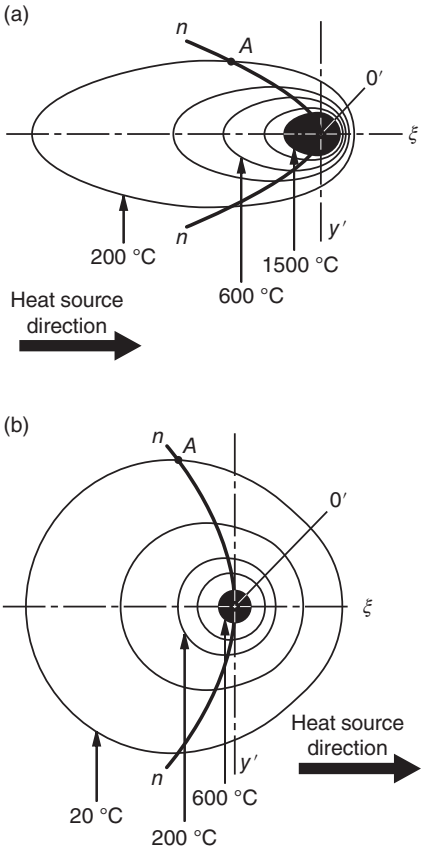


Figure 4.5 Temperature distribution in a plate for (a) low thermal conductivity material (low carbon steel) and (b) high thermal conductivity material (aluminum). Other processing conditions are the same. Source: From Rosenthal (1946)/American Society of Mechanical Engineers.

isotherms obtained for a thin plate and a thick plate, when the processing conditions are the same. These graphs were obtained by solving equations (4.17, 4.22). From these figures and the equations, the following deductions can be made:

1. The temperature gradient ahead of the heat source is much higher than that behind it.
2. Different points along the *y*-axis in a given section reach their peak temperature at different times. Points farther away have a lower peak temperature, and that is reached at a later time. The locus of points that reach their peak temperatures at the same instant is indicated by curve $n - n$ in Figure 4.5.
3. A higher thermal conductivity material such as aluminum makes the isotherms more circular, reducing the temperature gradient in front of the heat source (Figure 4.5).
4. Increasing the traverse speed makes the isotherms more elongated, while also increasing the lag of the locus $n - n$ (Figure 4.6).
5. Increasing the heat input or preheating does not change the shape of the isotherm but increases the size. This widens the fusion zone, as well as the heat-affected zone (HAZ) (Section 5.1.3).

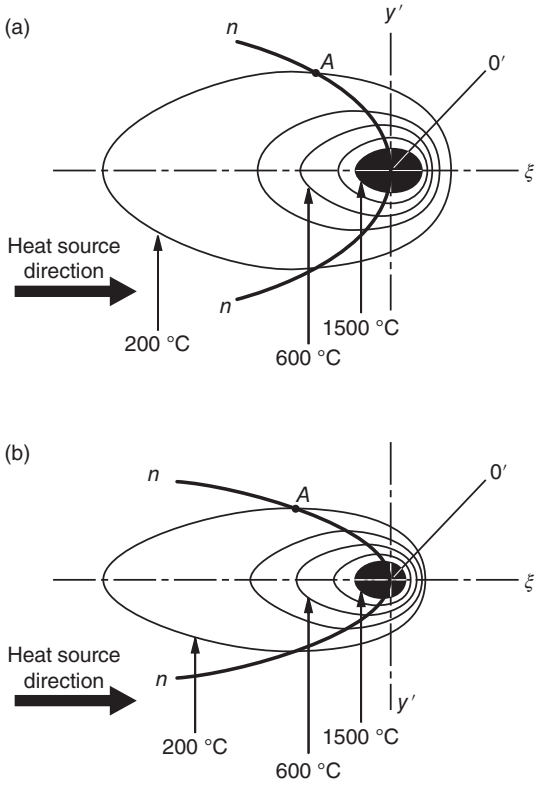


Figure 4.6 Temperature distribution as a function of processing speed, other processing conditions being the same. (a) Low speed. (b) High speed (Doubled). Source: From Rosenthal (1946)/American Society of Mechanical Engineers.

6. For the same conditions, a thinner plate results in a greater HAZ size than a thicker plate, while the thicker plate results in a higher temperature gradient (Figure 4.7).

Example 4.1 Two thin steel plates are welded using a laser beam. The following conditions are used:

Power input = 6 kW
Plate thickness = 2.5 mm
Welding speed = 50 mm/s
Initial plate temperature = 25 °C
Heat transfer efficiency = 0.7.
Calculate the temperature in the plate, 2.5 mm behind the laser beam and 2 mm to one side of it.

Solution:

From Appendices 4.D and 4.E, the properties of the steel plate are approximated as:

Average density, $\rho = 7870\,\text{kg/m}^3$
Average specific heat, $c_p = 452\,\text{J/kg K}$
Thermal conductivity, $k = 0.073\,\text{W/mm K}$.

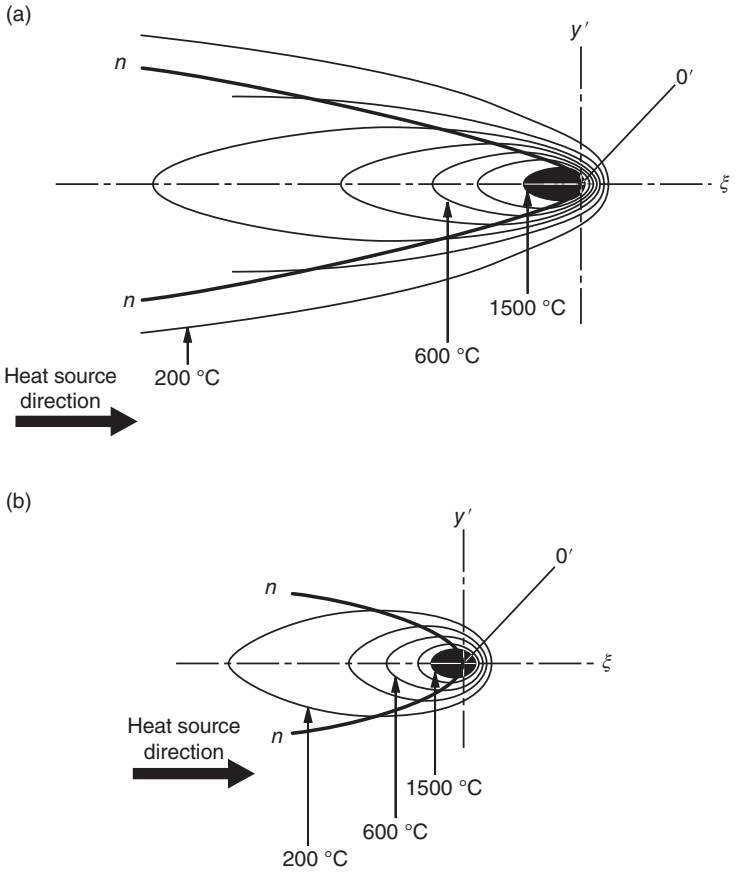


Figure 4.7 Temperature distribution as a function of plate thickness, other processing conditions being the same: (a) thin plate and (b) thick plate (more than 20 times thicker). Source: From Rosenthal (1946)/American Society of Mechanical Engineers.

The distance, r, of the point of interest from the heat source is given by:

$$r = \sqrt{\xi^2 + y^2} = \sqrt{(-2.5)^2 + 2^2} = 3.2 \text{ mm}$$

The value of ξ is negative, since it is behind the heat source and thus in the negative x-direction. The thermal diffusivity, κ, of the material is

$$\kappa = \frac{k}{\rho c_p} = \frac{0.073}{7.87 \times 10^{-6} \times 452} = 20.52 \text{ mm}^2/\text{s}$$

Thus,

$$\frac{u_x \xi}{2\kappa} = \frac{50 \times (-2.5)}{2 \times 20.52} = -3.045 \quad \text{and} \quad \frac{u_x r}{2\kappa} = \frac{50 \times 3.2}{2 \times 20.52} = 3.90$$

giving

$$K_0\left(\frac{u_x r}{2\kappa}\right) = K_0(3.90) = 0.01248$$

Now the power, q, available at the workpiece is

$$q = 0.7 \times 6000 = 4200 \text{ W}$$

and

$$T - T_0 = \frac{q}{2\pi kh} \exp\left(-\frac{u_x \xi}{2\kappa}\right) \times K_0\left(\frac{u_x r}{2\kappa}\right)$$

Therefore,

$$T = 25 + \frac{4200}{2\pi \times 0.073 \times 2.5} \times e^{-(-3.045)} \times 0.01248 = 985\,^{\circ}\text{C}$$

4.2.2 Peak Temperatures

The peak temperatures experienced throughout the workpiece determine the size of the HAZ. The peak temperature at a given point is experienced by the point shortly after it is passed by the heat source.

At any position of the heat source, the isotherms of various temperatures are oval-shaped. The point on any isotherm which is furthest from the x-axis (or line of motion of the heat source) is at its peak temperature at that instant. At these points, the tangent to the isotherm is parallel to the x-axis, and thus the following relation holds

$$\left(\frac{\partial y}{\partial x}\right)_T = 0 \tag{4.23}$$

Using Eqs. (4.17, 4.22, 4.23), and considering temperatures in terms of distance from the fusion zone boundary, it can be shown that the peak temperature for a thin plate (line source) is given (Problem 4.2) by:

$$\frac{1}{T_\text{p} - T_0} = \frac{\sqrt{2\pi e}\rho c_\text{p} h u_x Y}{q} + \frac{1}{T_\text{m} - T_0} \tag{4.24}$$

while that for a thick plate (point source) is

$$\frac{1}{T_\text{p} - T_0} = \frac{2\pi k \kappa e}{q u_x}\left[2 + \left(\frac{u_x Y}{2\kappa}\right)^2\right] + \frac{1}{T_\text{m} - T_0} \tag{4.25}$$

where e = natural exponent = 2.71828, T_p is the peak or maximum temperature at a distance Y from the fusion boundary (°C), and Y is the distance from the fusion boundary at the workpiece surface (mm).

Equations (4.24) and (4.25) are applicable to single-pass processes and have to be applied to each pass by itself. They are useful for estimating the HAZ size and also for showing the effect of preheat on the HAZ size. It is evident from the equations that all parameters being constant, preheating increases the size of the HAZ. Also, the size of the HAZ is proportional to the net energy input. Thus, high-intensity processes such as laser welding generally have a smaller HAZ. A high-intensity energy source results in a lower total heat input, because the energy used in melting the metal is concentrated in a small region.

In general, the equation (i.e. (4.24) or (4.25)) which gives the higher computed distance from the fusion zone or higher peak temperature at a given location is the more accurate of the two.

Example 4.2 A steel plate is welded using a laser of output power of 5 kW, at a speed of 40 mm/s, resulting in a heat transfer efficiency of 75%. The material has a thickness of 4 mm. If the ambient temperature is 30 °C, determine

(a) The peak temperature at a distance of 1.0 mm from the fusion boundary.
(b) The HAZ size.

Solution:
The properties of the steel plate are approximated as:
Average density, $\rho = 7870\,\text{kg/m}^3$
Average specific heat, $c_\text{p} = 452\,\text{J/kg K}$
Thermal conductivity, $k = 0.073\,\text{W/mm K}$
Melting temperature, $T_\text{m} = 1538\,^\circ\text{C}$.
With a heat transfer efficiency of 75%, the net power input available at the workpiece is

$$q = 0.75 \times 5 = 3.75\ \text{kW} = 3750\ \text{W}$$

The thermal diffusivity is given by:

$$\kappa = \frac{k}{\rho c_\text{p}} = \frac{0.073}{7.87 \times 10^{-6} \times 452} = 20.52\ \text{mm}^2/\text{s}$$

We calculate the peak temperature for both the two- and three-dimensional cases and select the higher value

(a) The peak temperature calculated at a distance of 1.0 mm from the fusion boundary, for the thin plate, is obtained from Eq. (4.24) as:

$$\frac{1}{T_\text{p} - 30} = \frac{\sqrt{2\pi e} \times 7.87 \times 10^{-6} \times 452 \times 4 \times 40 \times 1.0}{3750} + \frac{1}{1538 - 30}$$

giving

$$T_\text{p} = 805\,^\circ\text{C}$$

For the thick plate, we have from Eq. (4.25):

$$\frac{1}{T_p - 30} = \frac{2\pi \times 0.073 \times 20.52 \times 2.71828}{3750 \times 40}\left[2 + \left(\frac{40 \times 1.0}{2 \times 20.52}\right)^2\right] + \frac{1}{1538 - 30}$$

giving

$$T_\text{p} = 887.4\,^\circ\text{C}$$

Since the three-dimensional case results in the higher temperature, we consider the plate to be thick.

(b) In determining the HAZ size, the temperature at which the material will experience a change in properties will have to be defined. This temperature is used as the peak temperature in calculating the HAZ size. For example, for a plain carbon steel, the critical temperature is the eutectoid temperature. Thus, $T_p = 723\,^{\circ}\mathrm{C}$. Since we have determined that the thick plate formulation is more accurate, we now substitute this value in Eq. (4.25), giving

$$\frac{1}{723-30} = \frac{2\pi \times 0.073 \times 20.52 \times 2.71828}{3750 \times 40}\left[2 + \left(\frac{40 \times Y_{\mathrm{HAZ}}}{2 \times 20.52}\right)^2\right] + \frac{1}{1538-30}$$

Thus,

$$Y_{\mathrm{HAZ}} = 1.65\ \mathrm{mm}.$$

4.2.3 Cooling Rates

When a material is heated to a high enough temperature, the rate at which it cools afterward determines the grain structure and phases that are formed. These in turn affect mechanical properties such as strength and ductility. For example, high cooling rates result in a finer grain structure which increases the strength but reduces the ductility of the material. Knowledge of cooling rate is most important for materials which are polymorphic in nature, for example, steels. This enables a variety of phases with widely different mechanical properties to be produced. It is of less interest for aluminum, for example, where the cooling rates are always high.

In deriving an expression for the cooling rate, we simplify things by observing that the highest cooling rates occur along the line of motion of the heat source, that is, along the center line ($y = z = 0$, and $r = |\xi|$). We also consider only negative values of ξ, that is, behind the heat source. Substituting these conditions into Eq. (4.17) for the three-dimensional case (thick plate) gives

$$T - T_0 = \frac{q}{2\pi k \xi} \tag{4.26}$$

Differentiating with respect to t gives

$$\frac{dT}{dt} = \frac{\partial T}{\partial t} + \frac{\partial T}{\partial \xi}\frac{\partial \xi}{\partial t} \tag{4.27}$$

However, since the temperature undergoes no change with time with respect to the coordinate system attached to the heat source, the time derivative on the right-hand side of Eq. (4.27) vanishes ($\partial T(\xi, y, z, t)/\partial t = 0$), reducing Eq. (4.27) to the form:

$$\frac{dT}{dt} = \frac{\partial T}{\partial \xi}\frac{\partial \xi}{\partial t} \tag{4.28}$$

The first term of the product in Eq. (4.28) is obtained by differentiating Eq. (4.26) and substituting an equivalent expression for ξ from Eq. (4.26) as follows:

$$\frac{\partial T}{\partial \xi} = -\frac{q}{2\pi k \xi^2} = -\frac{q}{2\pi k}\left[\frac{2\pi k}{q}(T - T_0)\right]^2 = -\frac{2\pi k}{q}(T - T_0)^2 \tag{4.29}$$

Now using the chain rule, we have

$$\frac{\partial \xi}{\partial t} = \frac{\partial \xi}{\partial x}\frac{\partial x}{\partial t} = u_x$$

The centerline cooling rate then becomes

$$\frac{dT}{dt} = -\frac{2\pi k u_x}{q}(T - T_0)^2 \tag{4.30}$$

Eq. (4.30) is applicable to the three-dimensional case, that is, thick materials, and indicates that for such sizeable workpieces, the cooling rate is proportional to the square of the temperature rise above the initial temperature.

For thin plates, that is, the two-dimensional case, the centerline cooling rate can be derived in a similar fashion (Problem 4.5) and is given by

$$\frac{dT}{dt} = -2\pi k \rho c_p \left(\frac{u_x h}{q}\right)^2 (T - T_0)^3 \tag{4.31}$$

As indicated in Section 4.2.1, the plate is considered to be thin when $\beta_c < 0.6$ and thick when $\beta_c > 0.9$ (Eq. 4.9) or in general, the equation (i.e. (4.30) or (4.31)) which gives the lower cooling rate is the more accurate of the two (Figure 4.8). This is especially the case when $0.6 < \beta_c < 0.9$. It must be realized, however, that these methods are only approximate, and that where high accuracy is essential, numerical methods ought to be used.

From Eqs. (4.30) and (4.31), the following deductions can be made:

(a) Increasing the heat input (q) reduces the cooling rate, while increasing the traverse velocity (u_x) increases the cooling rate.

(b) Increasing the initial workpiece temperature (or preheat) (T_0) reduces the cooling rate and is more effective than increasing the heat input or reducing the traverse velocity.

(c) The cooling rate increases with an increase in plate thickness (h).

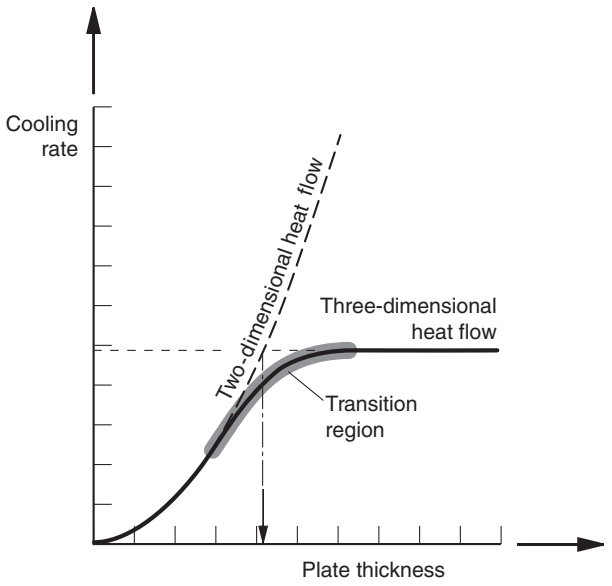


Figure 4.8 Representative regions for the cooling rate equations. Source: From Adams (1958)/Public Domain/U.S Departmet of Energy.

(d) A higher conductivity (k) material such as aluminum results in a higher cooling rate.
(e) The cooling rate decreases with an increasing distance (y) from the process center line.

Equations (4.30) and (4.31) strictly give the centerline cooling rates behind a point or line source of heat moving in a straight line at constant velocity on a flat surface and are most accurate for cooling rates at temperatures which are significantly below the melting temperature. Fortunately, the temperatures at which cooling rates are of metallurgical interest, especially for steels, are well below the melting point, and the estimates from these equations are then reasonably accurate. Furthermore, since the centerline cooling rate is only about 10% higher than in the HAZ, these equations also fairly well represent cooling rates in the regions of metallurgical interest.

Example 4.3 Calculate the cooling rate at 500 °C in a 3-mm-thick aluminum plate which is clad (Section 11.3) using a laser with output power of 4.5 kW and speed of 60 mm/s. Assume an ambient temperature of 25 °C and a heat transfer efficiency of 0.70. Use both the thin and thick plate equations and compare the results.

Solution:
From Appendices 4D and 4E, the properties of the aluminum plate are approximated as:
Average density, $\rho = 2700\,\text{kg/m}^3$
Average specific heat, $c_\text{p} = 900\,\text{J/kg K}$
Thermal conductivity, $k = 0.247\,\text{W/mm K}$.
The actual power available at the workpiece is

$$q = 0.70 \times 4500 = 3150\ \text{W}$$

First, we calculate β_c using Eq. (4.9) to determine whether the thick or thin plate equation is more appropriate:

$$\beta_\text{c} = h\sqrt{\frac{\rho c_\text{p} u_x (T - T_0)}{q}} = 3 \times \sqrt{\frac{2.7 \times 10^{-6} \times 900 \times 60 \times (500 - 25_0)}{3150}} = 0.44$$

Since $\beta_\text{c} < 0.6$, the plate may be considered thin. Thus, we have for the thin plate:

$$\frac{dT}{dt} = -2\pi \times 0.247 \times 2.7 \times 10^{-6} \times 900 \times \left(\frac{60 \times 3}{3150}\right)^2 \times (500 - 25)^3 = -1320\,^\circ\text{C/s}$$

For the thick plate solution, we have

$$\frac{dT}{dt} = \frac{-2\pi \times 0.247 \times 60}{3150} \times (500 - 25)^2 = -6670\,^\circ\text{C/s}$$

Since the cooling rate value calculated for the thin plate is the lower of the two, it is the more accurate one.

4.2.4 Gaussian Heat Source

The Gaussian distribution provides a better description of most heat sources than the point or line source described in preceding sections. It is especially important for processes such as heat treatment and laser forming, where the heat source is spread over a relatively wide area. However, it does not have a complete closed-form solution and requires numerical analysis. The basic elements of the analysis are presented in this section.

In analyzing the heat flow for a Gaussian heat source, the workpiece is assumed to be semi-infinite (three-dimensional), that is, $-\infty < x, y < \infty$ and $0 \le z < \infty$. The heat supply in the material per unit volume and unit time at a point (x', y', z') due to a Gaussian heat distribution for a constant power source q is

$$q_{s1}(x', y', z') = \frac{\alpha q(1-R)}{2\pi w^2} \times e^{-\left[\frac{x'^2+y'^2}{2w^2}+\alpha z'\right]} \quad \text{for} \quad 0 < z' < \infty \tag{4.32}$$

where w is the Gaussian beam radius and α is the beam power absorption coefficient which is a property of the base material. It is about $800\,\mathrm{m}^{-1}$ for mild steel in keyhole welding and R is the reflectivity, which is about 0.3 for steel.

The initial conditions of the problem are given by:

$$T(x, y, z, 0) = 0 \quad \text{everywhere in the domain} \tag{4.33}$$

The boundary conditions are (for $t > 0$)

$$T(x, y, z, t) = 0 \quad \text{for} \quad \begin{cases} x, y \to \pm\infty \\ z \to \infty \end{cases} \tag{4.34}$$

Convection and radiation heat losses at the top surface of the workpiece are neglected. Thus, the boundary condition at the work surface, $z = 0$, is

$$-k\frac{\partial T(x, y, z, t)}{\partial z} = 0 \quad \text{at} \quad z = 0 \tag{4.35}$$

The last boundary condition can be satisfied by using the imaginary heat source technique. For any real point heat source in the region (x', y', z'), we assume that there is a corresponding imaginary heat source at $(x', y', -z')$ of the same strength as the real source. We also assume an infinite medium instead of a semi-infinite one. This makes the temperature field symmetric about the $z = 0$ plane, that is, there is no heat transfer across the work surface ($z = 0$). Eq. (4.34) then becomes

$$T(x, y, z, t) = 0 \quad \text{for} \quad x, y, z \to \pm\infty \tag{4.36}$$

Considering the real and imaginary heat sources, q_{s1} in Eq. (4.32) becomes

$$q_{s1}(x', y', z') = \frac{\alpha q(1-R)}{2\pi w^2} \times e^{-\left[\frac{x'^2+y'^2}{2w^2}+\alpha|z'|\right]} \quad \text{for} \quad -\infty < z' < \infty \tag{4.37}$$

where $|z'|$ is the absolute value of z' so that q_{s1} is symmetric about the $z = 0$ plane.

Now suppose that at time $t = t'$, a unit heat pulse is emitted at (x', y', z'), and that an infinite medium moves uniformly past (x', y', z') with velocity $\mathbf{u} = (u_x, 0, 0)$, $u_x < 0$, parallel to the x-axis.

The gage temperature at the fixed point (x, y, z) and time t due to the unit heat pulse emitted at (x', y', z') and time t' is

$$G_r(x,y,z,t) = \frac{1}{8\rho c_p[\pi\kappa(t-t')]^{3/2}} \times e^{-\left[\frac{[x-u_x(t-t')-x']^2+(y-y')^2+(z-z')^2}{4\kappa(t-t')}\right]} \tag{4.38}$$

Thus, for a Gaussian source of power q, and a medium moving with velocity $\mathbf{u}$, the temperature rise over the period $t' = 0$ to t over the body is

$$\begin{aligned} T(x,y,z,t) &= \int_0^t \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} q_{s1}(x',y',z') \times G_r(x',y',z',t' \mid x,y,z,t)dx'dy'dz'dt' \\ &= \int_0^t \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\alpha q(1-R)}{2\pi w^2} \times e^{-\left[\frac{x'^2+y'^2}{2w^2}+\alpha|z'|\right]} \times \frac{1}{8\rho c_p[\pi\kappa(t-t')]^{3/2}} \\ &\quad \times e^{-\left[\frac{[x-u_x(t-t')-x']^2+(y-y')^2+(z-z')^2}{4\kappa(t-t')}\right]} dx'dy'dz'dt' \end{aligned} \tag{4.39}$$

This reduces to

$$\begin{aligned} T(x,y,z,t) = \int_0^t \frac{\alpha q(1-R)}{4\pi\rho c_p[w^2+2\kappa(t-t')]} \times e^{\left[\kappa\alpha^2(t-t')-\frac{[x-u_x(t-t')]^2+y^2}{4\kappa(t-t')+2w^2}\right]} \\ \times \left\{ e^{\alpha z} erfc\left(\alpha\sqrt{\kappa(t-t')} + \frac{z}{2\sqrt{\kappa(t-t')}}\right) + e^{-\alpha z} erfc\left(\alpha\sqrt{\kappa(t-t')} - \frac{z}{2\sqrt{\kappa(t-t')}}\right)\right\} dt' \end{aligned} \tag{4.40}$$

where erfc (x) is the complementary error function, defined by:

$$erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-\varsigma^2} d\varsigma$$

Equation (4.40) does not have a closed-form solution, and thus it is necessary to use numerical integration to obtain the distribution at time t.

4.2.5 The Two-Temperature Model

The two-temperature model to be discussed in this section is more appropriate for describing the interaction between ultrashort pulse beams and metals. The mechanism of laser–material interaction for ultrashort pulse lasers (sub-picosecond pulse duration) is different from that for the traditional long pulse lasers (nanosecond and above). Whereas the electrons and lattice are essentially in thermal equilibrium at the end of a long pulse, that is not necessarily the case for ultrashort pulse beams. For the latter case, electrons and the metal lattice (phonons) behave as two separate systems. Electrons located within the skin depth first absorb the photon energy. This is then transferred to the metal lattice through electron–phonon collisions. Even though the same process occurs for both long and ultrashort pulse lasers, their separate effects are more evident at short interaction times. As the electrons absorb the photon energy, they get excited to higher energy levels, and rapidly thermalize, resulting in a hot free electron gas. The high-temperature electron gas then diffuses into the

metal, heating up the metal lattice through electron–phonon collisions. The collision or thermalization time of the electrons is of the order of 20 femtoseconds at room temperature. However, the electron–phonon thermal relaxation time is of the order of picoseconds, since several collisions are necessary for effective energy exchange between electrons and phonons.

Local thermal equilibrium is thus established between the hot electrons and the lattice during the pulse when the laser pulse duration is longer than the electron–phonon thermal relaxation time. This results in the electrons and lattice having the same temperature by the end of the pulse. The analyses in the preceding sections are then appropriate. These constitute the one-step heating model.

Electrons can normally be heated to very high temperatures within short time periods, since their heat capacity is relatively low. Thus for situations where the pulse duration is of the same order of magnitude as the electron–phonon thermal relaxation time or less, the electrons and lattice are no more in local thermal equilibrium. They then have to be considered as separate systems and modeled as a two-step process that constitutes

(a) Absorption of photon energy by electrons.
(b) Lattice heating through electron–phonon coupling.

The corresponding basic governing equations for the two-step process are as follows:

$$c_{\text{vole}}(T_{\text{e}})\frac{\partial T_{\text{e}}}{\partial t} = \nabla \cdot (k_{\text{e}} \nabla T_{\text{e}}) - \Delta Q_{\text{e}\to\text{l}} + q_{\text{sl}} \tag{4.41}$$

$$c_{\text{voll}}(T_l)\frac{\partial T_l}{\partial t} = \Delta Q_{e\to l} \tag{4.42}$$

where $c_{\text{vol}} = \rho c_{\text{p}}$ is the heat capacity (specific heat per unit volume), the subscripts e and l refer to electron and lattice, respectively, and q_{sl} is the volumetric heat source term given by Eq. (4.32) for the three-dimensional case.

$\Delta Q_{\text{e}\to\text{l}}$ is the energy transferred by electrons to the lattice and can be approximated from free electron theory (provided the lattice temperature is not much smaller than the Debye temperature) as:

$$\Delta Q_{\text{e}\to\text{l}} = \frac{\pi^2 m_{\text{e}} n_{\text{e}} {u_{\text{s}}}^2}{6 t_{\text{c}}(T_{\text{e}}) T_{\text{e}}}(T_{\text{e}} - T_{\text{i}}) = G(T_{\text{e}} - T_{\text{i}}) \tag{4.43}$$

where m_{e} is the electron mass, n_{e} is the electron number density (/m^3), u_{s} is the speed of sound, and $t_{\text{c}}(T_{\text{e}})$ is the electron mean free time between collisions at temperature T_{e}.

This basic form of the two-step model is sometimes referred to as the parabolic two-step model (PTS), while the traditional model is referred to as the parabolic one-step model (POS). Since the expression for G in Eq. (4.43) can be shown to be only weakly dependent on T_{e}, it can be further simplified as:

$$G = \frac{\pi^4 (n_{\text{e}} u_{\text{s}} k_{\text{B}})^2}{18k} \tag{4.44}$$

Sample measured values for G are shown in Table 4.1.

This formulation of the governing equations neglects the energy transport by phonons, since heat flux is carried mainly by free electrons for metals. The variation of the electron and lattice temperatures with time on an irradiated surface are illustrated in Figure 4.9a. The surface electron temperature is much higher than the lattice temperature in the initial stages of the process as the electrons absorb the incident radiation. However, the two temperatures eventually become equal after a long enough time as thermalization occurs. The lattice temperature profiles obtained using the

Table 4.1 Measured values of the electron–phonon coupling factor, *G*, for some metals (units: 10^{16} W/(m^3 K)).

Metal	G	Metal	G
Copper (Cu)	4.8±0.7	Vanadium (V)	523±37
Silver (Ag)	2.8	Niobium (Nb)	387±36
Gold (Au)	2.8±0.5	Titanium (Ti)	185±16
Chromium (Cr)	42±5	Lead (Pb)	12.4±1.4
Tungsten (W)	26±3		

Source: Qiu and Tien (1992)/with permission of Elsevier.

two-temperature model are compared with those obtained for the corresponding one-temperature or traditional model in Figure 4.9b for a 10 ps laser pulse. The traditional model tends to confine the absorbed radiation energy mainly in regions close to the surface, resulting in much higher peak temperatures being predicted, while the two-temperature model predicts a much larger HAZ. Similar predictions for a nanosecond laser pulse (Figure 4.9c) shows little difference between the two models, indicating that the effects of the microscopic radiation–metal (photon–electron and electron–phonon) interactions are then less important.

The two-temperature model can be extended to account for energy transport in the hyperbolic formulation. This is referred to as the hyperbolic two-step model (HTS) when only energy transport by electrons is considered. When transport in the lattice is also considered, then we have the dual HTS model, which is given by the following equations:

$$c_{\text{vole}}(T_\text{e})\frac{\partial T_\text{e}}{\partial t} = -\nabla \cdot \mathbf{q}_\text{e} - G(T_\text{e} - T_\text{l}) + q_\text{sl} \tag{4.45}$$

$$\tau_\text{Fe}\frac{\partial q_\text{e}}{\partial t} + q_\text{e} = -k_\text{e}\nabla T_\text{e} \tag{4.46}$$

$$c_{\text{voll}}(T_\text{l})\frac{\partial T_\text{l}}{\partial t} = -\nabla \cdot \mathbf{q}_\text{l} + G(T_\text{e} - T_\text{l}) \tag{4.47}$$

$$\tau_\text{Fl}\frac{\partial q_\text{l}}{\partial t} + q_\text{l} = -k_\text{l}\nabla T_\text{l} \tag{4.48}$$

where $\mathbf{q}_\text{e}$ and $\mathbf{q}_\text{l}$ are the heat flux vectors for the electrons and lattice, respectively, and τ_F is the relaxation time evaluated at the Fermi surface. The relaxation time is the mean time for electrons to change their states.

Now the thermalization time, τ_t, that is, characteristic time for electrons and phonons to reach equilibrium, can be expressed as:

$$\tau_\text{t} = \frac{c_{\text{vole0}}(T)}{G} \tag{4.49}$$

where the subscript 0 indicates evaluation at a reference temperature.

The hyperbolic transport effect on the electron temperature is more important when the laser pulse duration is shorter than τ_F, and the mechanism of energy deposition becomes important when the pulse duration is shorter than τ_t. Values of τ_Fe and τ_t for some metals are shown in Table 4.2. At room temperature, τ_Fe is about an order of magnitude smaller than τ_t. Thus, the mechanisms of energy deposition are expected to play much more important roles during short pulse laser heating of metals than the effects of hyperbolic transport.

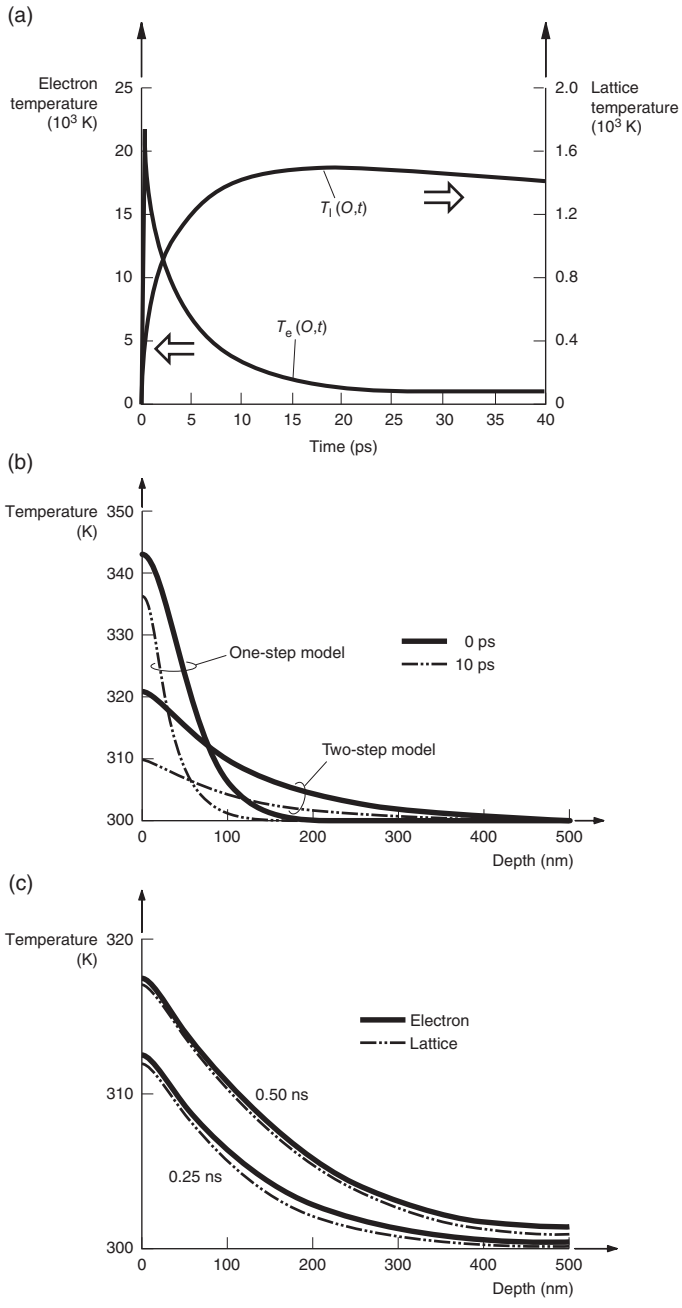


Figure 4.9 (a) Electron and lattice temperature variations with time on the surface of a gold film of thickness 200 nm heated with a 40 fs laser pulse of fluence 0.5 J/cm^2. (b) Lattice temperature profiles for a gold film of thickness 0.5 μm heated with a 10 ps laser pulse of fluence 10 mJ/cm^2. (c) Electron and lattice temperature profiles for a gold film of thickness 1 μm heated with a 1 ns laser pulse of fluence 20 mJ/cm^2. Source: (a) Adapted from Chen et al. (2005). (b) Adapted from Qiu and Tien (1993).

Table 4.2 Thermalization τ_t and electron relaxation τ_{Fe} times at 300 K for some metals (units: ps).

Metal	τ_{Fe}	τ_t	Metal	τ_{Fe}	τ_t
Copper (Cu)	0.03	0.6	Vanadium (V)	0.002	0.06
Silver (Ag)	0.04	0.6	Niobium (Nb)	0.004	0.05
Gold (Au)	0.04	0.8	Titanium (Ti)	0.001	0.05
Chromium (Cr)	0.003	0.1	Lead (Pb)	0.005	0.4
Tungsten (W)	0.01	0.2			

Source: Adapted from Qiu and Tien (1993).

The PTS and POS models are special cases of the HTS model. The HTS model becomes the PTS model when the laser pulse duration is much longer than τ_{Fe}, and it becomes the POS model when the pulse duration is much longer than both τ_{Fe} and $\tau_{t.}$

For one-dimensional analysis, if the spatial distribution of the laser beam is considered uniform, and the temporal shape of the pulse is assumed to be Gaussian with a full width at half maximum (FWHM) pulse duration of τ_p, then q_{sl} in Eq. (4.41) can be expressed as:

$$q_{sl} = \alpha \frac{1}{\tau_p} \frac{Q(1-R)}{\pi w^2} \times e^{\left[-\alpha z - 4\ln 2\left(\frac{t}{\tau_p}\right)^2\right]} \tag{4.50}$$

where Q is the pulse energy, J.

In the next section, we address the basic concepts associated with flow in the molten pool. The discussion is useful for processes that involve melting, such as surface melting and welding.

4.3 FLUID FLOW IN MOLTEN POOL

Fluid flow in the molten pool affects the structure and properties of the fused zone, such as the geometry (penetration and bead width), surface smoothness (ripples), porosity, lack of fusion, undercutting, solidification structure, and macrosegregation (mixing). The principal flow factors that determine the process quality include the flow velocities and mode of solidification for the processes that involve melting. In this section, we focus on flow in the molten pool. Solidification is discussed in the next two chapters.

The flow system is illustrated schematically in Figure 4.10 for conduction mode laser welding (or melting) (Sections 10.3.1 and 11.2). The primary forces responsible for flow in the molten pool for this case are the surface tension and buoyancy forces. The equations governing convection and fluid flow in the molten pool are obtained by considering mass conservation in the pool (continuity equation), momentum due to motion of the fluid (Navier–Stokes equations), and conduction in the workpiece. Heat conduction in the workpiece has already been discussed in the preceding section. In the following subsections, we discuss the continuity and momentum equations that govern flow in the molten pool. This is followed by a discussion on the surface tension effect for a flat pool surface. Finally, the basic concepts involved in analyzing a pool with a free surface are outlined.

4.3.1 Continuity Equation

Using the same coordinate system as shown in Figures 4.2, 4.3, and 4.10, the general form of the differential equation for conservation of mass in a fluid medium is expressed as:

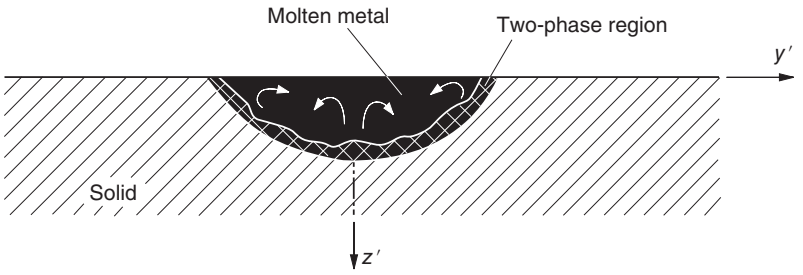


Figure 4.10 Configuration of the molten pool.

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho u_x)}{\partial x} + \frac{\partial(\rho u_y)}{\partial y} + \frac{\partial(\rho u_z)}{\partial z} = 0 \quad \text{or} \quad \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0 \tag{4.51}$$

For a steady flow and incompressible fluid, the density is constant, and thus the equation reduces to the form:

$$\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} + \frac{\partial u_z}{\partial z} = 0 \quad \text{or} \quad \nabla \cdot \mathbf{u} = 0 \tag{4.52}$$

where $\mathbf{u}$ is the velocity vector, m/s and u_i = velocity components, m/s.

4.3.2 Navier–Stokes Equations

Motion or flow of the fluid in the molten pool is also governed by the Navier–Stokes equations. Assuming incompressible fluid with negligible variations in the viscosity, these become

$$\rho \frac{d\mathbf{u}}{dt} = -\nabla P + \mu \nabla^2 \mathbf{u} + \mathbf{F}_b \quad \text{or} \quad \rho \left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} \right) = -\nabla P + \mu \nabla^2 \mathbf{u} + \mathbf{F}_b \tag{4.53}$$

which can be expanded to the form:

$$\rho \left(\frac{\partial u_x}{\partial t} + u_x \frac{\partial u_x}{\partial x} + u_y \frac{\partial u_x}{\partial y} + u_z \frac{\partial u_x}{\partial z} \right) = -\frac{\partial P}{\partial x} + \mu \left(\frac{\partial^2 u_x}{\partial x^2} + \frac{\partial^2 u_x}{\partial y^2} + \frac{\partial^2 u_x}{\partial z^2} \right) + F_x \tag{4.54a}$$

$$\rho \left(\frac{\partial u_y}{\partial t} + u_x \frac{\partial u_y}{\partial x} + u_y \frac{\partial u_y}{\partial y} + u_z \frac{\partial u_y}{\partial z} \right) = -\frac{\partial P}{\partial y} + \mu \left(\frac{\partial^2 u_y}{\partial x^2} + \frac{\partial^2 u_y}{\partial y^2} + \frac{\partial^2 u_y}{\partial z^2} \right) + F_y \tag{4.54b}$$

$$\rho \left(\frac{\partial u_z}{\partial t} + u_x \frac{\partial u_z}{\partial x} + u_y \frac{\partial u_z}{\partial y} + u_z \frac{\partial u_z}{\partial z} \right) = -\frac{\partial P}{\partial z} + \mu \left(\frac{\partial^2 u_z}{\partial x^2} + \frac{\partial^2 u_z}{\partial y^2} + \frac{\partial^2 u_z}{\partial z^2} \right) + F_z \tag{4.54c}$$

where $\boldsymbol{F}_b$ is the body force vector with components F_x, F_y, F_z (N/m^3), P is the local fluid pressure (Pa), and μ is the dynamic or absolute viscosity (Pa s).

In laser processing, Maxwell's equations do not play a role, since there are no electromagnetic forces. Thus, the only body force left is the gravitational force in the z-direction. Hence the body force, using the Boussinesq approximation, for the configuration shown in Figure 4.10, reduces to

$$F_z = -\rho g \beta_T (T - T_o) \quad F_x = F_y = 0$$

where g is the gravitational acceleration (m/s^2) and β_T is the volumetric thermal coefficient of expansion (/K).

Here, the pressure, P, is defined to include the hydrostatic component, ρgh. Equations (4.51) to (4.54), along with Eq. (4.2a,b) constitute the governing equations for flow in the molten pool when the conduction mode is predominant (i.e. there is no keyhole (Section 10.3.1)). They are coupled since the buoyancy term and surface tension forces provide a link between the flow and thermal equations. With the appropriate boundary conditions, they can be solved to obtain the temperature and velocity fields in the molten pool. These equations are more appropriate for processes such as conduction mode welding and surface melting. Fluid flow in laser cutting and keyhole welding will be discussed under the appropriate topics (Chapters 9 and 10). In the following section, the influence of surface tension on flow in the molten pool is further discussed.

4.3.3 Surface Tension Effect

The surface tension that exists on the molten pool surface induces thermocapillary (surface tension driven) flow on the pool surface. This is because the surface tension force changes with temperature. And since the temperature varies on the pool surface, a surface tension gradient exists on the surface. This results in fluid being drawn along the surface from the region of lower surface tension to the region of higher surface tension. The flow induced by surface tension is referred to as Marangoni convection. The bead shape that results depends on whether the temperature derivative of the surface tension (surface tension gradient) $\frac{d\gamma}{dT}$ is positive or negative (Figure 4.11).

The form of the surface tension gradient is determined by surface-active elements (i.e. elements that preferentially segregate to the surface of the molten metal) such as sulfur and oxygen (for iron-based alloys) in the base material. These elements alter both the surface tension and temperature dependence of the surface tension. Thus, different heats of the same material (compositions within the specified allowable range) are known to produce different shapes of the fusion zone under the same processing conditions. This is because small variations in the sulfur content, for example, result in considerable variations in penetration.

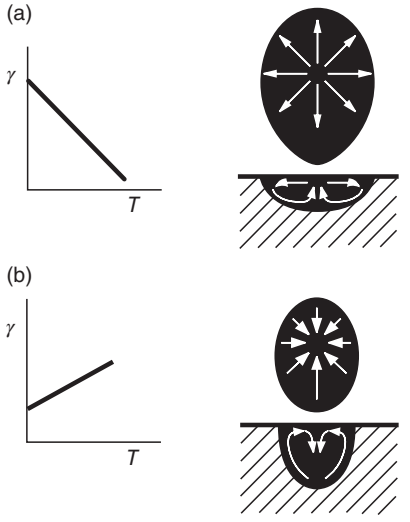


Figure 4.11 (a) Flow pattern for a negative gradient. (b) Flow pattern for a positive gradient. Source: Adapted from Heiple and Roper (1981).

Ordinarily, metals that are either pure or that have insignificant amounts of surface-active elements have a negative surface tension gradient and thus induce outward flow of molten metal on the pool surface. Hotter material from the center is thus carried outward toward the periphery at high velocities. As the flow approaches the edge of the molten pool, it is diverted downward and then turned around, moving back to the pool center to complete the recirculation pattern (Figure 4.11a). This results in a wider but shallower pool. The surface flow normally occurs in a thin layer called the Marangoni layer.

For a molten pool with a positive surface tension gradient, the surface tension forces will increase with temperature, that is, from the periphery or toe (which is relatively cooler) toward the center of the molten pool (which is relatively hotter). The higher surface tension forces in the center will draw or pull the material at the pool surface from the periphery toward the center. This results in surface flow inward (Figure 4.11b). Since the bead shape is ideally symmetrical, the stream of metal from the two edges of the pool will meet and flow downward at the center, thereby carrying the hottest material from the central portion of the free surface to the bottom of the pool, resulting in deeper penetration.

Sulfur, oxygen, and sellenium are among the elements that produce a positive surface tension temperature coefficient. As an illustration, addition of about 50 ppm of sulfur to a steel is adequate to increase the depth of penetration by over 80%, while about 140 ppm of selenium will increase the depth to width ratio by more than 160%.

Addition of surface-active elements to the molten pool can be accomplished by doping the filler rod, pre-coating the joint before processing, or doping the shielding gas.

The surface tension effect is specified as boundary shear at the free surface, that is, boundary conditions on the pool surface, since tangential force balance requires that the thermocapillary stresses should be balanced by viscous shear. The equations for the boundary shear (Eq. (4.55)) can be obtained by considering the fact that for Newtonian flow, the shear stress in the material (right-hand side) is proportional to the resulting shear strain rate (left-hand side). For the simple molten pool shown schematically in Figure 4.10, we obtain these equations by making the following assumptions:

1. Shear in the shielding gas immediately above the pool surface is negligible.
2. The pool surface is flat.

Then, we have

$$\mu \frac{\partial u_x}{\partial z} = -\frac{\partial T}{\partial x}\frac{\partial \gamma}{\partial T} \tag{4.55a}$$

$$\mu \frac{\partial u_y}{\partial z} = -\frac{\partial T}{\partial y}\frac{\partial \gamma}{\partial T} \tag{4.55b}$$

$$u_z = 0 \tag{4.55c}$$

where γ is the surface tension of the molten pool material at temperature T (J/m^2).

Surface tension dominated flow will normally result in maximum surface velocity of the order of 10^3 mm/s, while for buoyancy dominated flow, maximum surface velocities are of the order of 10 mm/s. This indicates that the thermocapillary forces are much greater than the buoyancy forces.

When a keyhole is formed, the depth to width ratio is not significantly influenced by the relatively small amounts of surface-active elements normally present in steels. This is because the penetration

under such conditions is affected primarily by the magnitude of surface tension rather than the surface tension gradient.

4.3.4 Free Surface Modeling

In the previous section, the pool surface was assumed to be flat. In reality, the molten pool surface undergoes both spatial and temporal variations during processing. As a result of the surface curvature, the surface tension stresses have both tangential and normal components. The surface stresses may be expressed in the general form:

$$\left[P^{(2)} - P^{(1)} + \frac{\gamma}{r_s}\right] n_i = \left[\mu^{(2)}\left(\frac{\partial u_i}{\partial x_k}^{(2)} + \frac{\partial u_k}{\partial x_i}^{(2)}\right) - \mu^{(1)}\left(\frac{\partial u_i}{\partial x_k}^{(1)} + \frac{\partial u_k}{\partial x_i}^{(1)}\right)\right] n_k - \frac{\partial \gamma}{\partial x_i} \tag{4.56}$$

where $P^{(1,2)}$, $\mu^{(1,2)}$, ${u_i}^{(1,2)}$, and ${u_k}^{(1,2)}$ are the pressures, dynamic viscosities, and velocity components in phases 1 and 2, respectively, r_s is the radius of curvature of the surface (m), and n_i ($i = 1, 2, 3$) are the components of the unit vector normal to the surface and directed into the interior of phase 1. Summation over a repeated index ($i, k = 1, 2, 3$) is assumed.

In two-dimensional problems, the radius of curvature, r_s, is given by

$$\frac{1}{r_s} = \frac{\frac{\partial^2 H}{\partial x^2}}{\left[1 + \left(\frac{\partial H}{\partial x}\right)^2\right]^{1.5}} \tag{4.57}$$

For three-dimensional problems, $1/r_s$ is replaced by:

$$\frac{1}{r_{sx}} + \frac{1}{r_{sy}} = \frac{\frac{\partial^2 H}{\partial x^2} + \frac{\partial^2 H}{\partial y^2}}{\left[1 + \left(\frac{\partial H}{\partial x}\right)^2 + \left(\frac{\partial H}{\partial y}\right)^2\right]^{1.5}} \tag{4.58}$$

where H is the height of the surface relative to some reference elevation.

Taking the projections of Eq. (4.56) in the directions normal and tangential to the interphase boundary, we obtain the curvature and Marangoni effects, respectively. The curvature effects (surface tension effects) are induced from normal stress balance at the free surface and can be represented as:

$$P^{(2)} - P^{(1)} + \frac{\gamma}{r_s} = 2\mu^{(2)}\frac{\partial u_n}{\partial n}^{(2)} - 2\mu^{(1)}\frac{\partial u_n}{\partial n}^{(1)} \tag{4.59}$$

where n indicates the direction normal to the interface.

The shear stress also balances the surface tension gradient induced by the temperature variation. This Marangoni effect is expressed as:

$$\mu^{(2)}\left(\frac{\partial u_n}{\partial \tau}^{(2)} + \frac{\partial u_\tau}{\partial n}^{(2)}\right) - \mu^{(1)}\left(\frac{\partial u_n}{\partial \tau}^{(1)} + \frac{\partial u_\tau}{\partial n}^{(1)}\right) = \frac{\partial \gamma}{\partial \tau} \tag{4.60}$$

where τ indicates the tangential direction of the free surface.

Because the fluid particle at the free surface must remain attached to the fluid surface, a kinematic condition needs to be satisfied and can be represented as:

$$\frac{\partial H}{\partial t} + u_x \frac{\partial H}{\partial x} + u_y \frac{\partial H}{\partial y} = u_z \tag{4.61}$$

4.4 SUMMARY

This chapter addressed basic concepts of heat and fluid flow as they apply to laser processing and was broken down into three main sections:

- Energy balance.
- Heat flow in a solid workpiece.
- Fluid flow in a molten pool.

The analysis on heat flow in a solid workpiece is directly applicable to processes such as laser forming and heat treatment, where no melting occurs. It also provides reasonable approximation for the heat flow in the solid part of the other processes that involve melting. With the heat source being considered as either a point or a line source, the analysis indicates that the temperature gradient ahead of the heat source is much higher than that behind it; increasing the traverse speed elongates the isotherms surrounding the heat source; higher thermal conductivity materials make the isotherms more circular, reducing the temperature gradient in front of the heat source; increasing the heat input or preheating does not change the shape of the isotherm but increases the size, thereby widening the fusion zone and HAZ and that for the same conditions, a thinner plate results in a greater HAZ size than a thicker plate.

The peak temperatures experienced throughout the workpiece determine the size of the HAZ. The peak temperature at a given point is experienced by the point shortly after it is passed by the heat source. The size of the HAZ increases with both preheating and the net energy input. Thus, the high-intensity laser processes generally have a smaller HAZ.

The cooling rate experienced by a material determines the grain structure and phases that are formed. Increasing the heat input reduces the cooling rate, while increasing the traverse velocity increases the cooling rate. Preheating reduces the cooling rate, while increasing the plate thickness increases the cooling rate. Higher conductivity materials such as aluminum result in a higher cooling rate.

A Gaussian distribution, which is more representative of the heat source, was then considered. This was followed by the two-temperature model, which is appropriate for ultrashort pulse laser beams in metals. For the latter case, the electron and lattice heat flows are modeled separately, with a coupling term. From this analysis, the surface electron temperature is found to be much higher than the lattice temperature in the initial stages of the process as the electrons absorb the incident radiation. However, the two temperatures become equal as thermalization occurs.

The analysis in the third section on flow in the molten pool is applicable to processes that involve melting such as welding, cladding, surface melting, and cutting. It was noted that a negative surface tension gradient (typical of pure metals) induces outward flow of molten metal on the pool surface, resulting in a wider and shallower pool. A positive surface tension gradient (which occurs when surface-active elements such as sulfur are present), on the other hand, induces inward flow of molten metal, resulting in deeper penetration for processes that do not involve a keyhole.

APPENDIX 4.A

List of symbols used in the chapter

Symbol	Parameter	Units
Subscripts e, l	Refer to electron and lattice, respectively	—
$\boldsymbol{F}_b$	Body force vector (F_x, F_y, F_z)	N/m^3
G	Electron–phonon coupling factor	W/(m^3 K)
H	Height of a surface relative to a reference elevation	mm
k_l	Constant	m K
q_e	Heat flux vector for electrons	W/mm^2
q_l	Heat flux vector for the lattice	W/mm^2
q_n	Net rate of heat flux per unit volume across element faces	W/mm^3
q_s	Rate of local internal energy generated per unit volume	W/mm^3
q_{sl}	Heat supply per unit volume by the power input q	W/mm^3
q_v	Rate of change of internal energy per unit volume	W/mm^3
$\Delta Q_{e\to l}$	Energy transferred by electrons to the lattice	J
T_{es}	Electron surface temperature	°C
Y	Distance from the fusion boundary at the workpiece surface	mm
$\frac{\partial T}{\partial t}$	Rate of change of temperature in moving coordinate system	°C/s
β_c	Constant for comparing two- and three-dimensional heat flow	—
$\phi\,(\xi, y, z)$	Analytic function to be determined in relation to temperature	°C
τ	Tangential direction of a free surface	—
$\tau_c(T_e)$	Electron mean free time between collisions at temperature T_e	s
τ_F	Relaxation time evaluated at the Fermi temperature	s
τ_t	Thermalization time	s
χ	argument of the Bessel function	—

APPENDIX 4.B DERIVATION OF EQUATION (4.2a)

In deriving Eq. (4.2a), use is made of Fourier's law for heat conduction across a surface, which states that

$$q = -kA\frac{dT}{dx} \tag{4.B.1}$$

where A is the area perpendicular to the flow of heat and q is the heat flux (rate of heat flow) across a surface.

To obtain the temperature distribution as a function of time, we consider a control volume as shown in Figure 4B.1. Considering the principle of energy conservation, the rate of change of internal energy per unit volume (q_v) of the control volume must equal the sum of:

1. Net rate of heat flux per unit volume across its faces, q_n.
2. Any heat sources or sinks within it per unit volume, q_s, such as from chemical reactions or current passing through it (Joule effect):

$$q_v = q_n + q_s \tag{4.B.2}$$

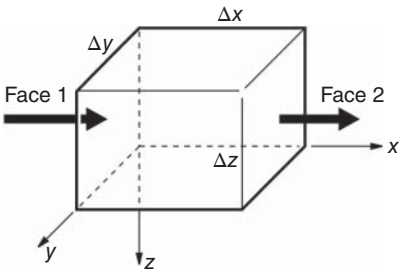


Figure 4B.1 Heat flux in a control volume.

The rate of change of internal energy per unit volume q_v is given by:

$$q_v = \rho c_p \frac{dT}{dt} \tag{4.B.3}$$

To obtain the net rate of heat flow per unit volume across the faces, q_n, we consider the heat flux across the individual faces. For face 1,

$$q_1 = -\left(k\Delta y\Delta z\frac{\partial T}{\partial x}\right)_x \tag{4.B.4a}$$

and across face 2:

$$q_2 = -\left(k\Delta y\Delta z\frac{\partial T}{\partial x}\right)_{x+\Delta x} \tag{4.B.4b}$$

Thus, the net rate of heat transfer per unit volume in the x-direction into the control volume is

$$\frac{q_1 - q_2}{\Delta x\Delta y\Delta z} = \frac{\left(k\frac{\partial T}{\partial x}\right)_{x+\Delta x} - \left(k\frac{\partial T}{\partial x}\right)_x}{\Delta x} \tag{4.B.5}$$

In the limit, as $\Delta x \to 0$, we have

$$\lim_{\Delta x\to 0}\frac{q_1 - q_2}{\Delta x\Delta y\Delta z} = \frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right) \tag{4.B.6}$$

Similar expressions hold for the y- and z-directions. Thus considering all three directions, we have

$$q_n = \frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial y}\left(k\frac{\partial T}{\partial y}\right) + \frac{\partial}{\partial z}\left(k\frac{\partial T}{\partial z}\right)$$

Then, from Eq. (4B.2) and (4B.3), we get

$$\rho c_p\frac{dT}{dt} = \frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial y}\left(k\frac{\partial T}{\partial y}\right) + \frac{\partial}{\partial z}\left(k\frac{\partial T}{\partial z}\right) + q_s \tag{4.B.7a}$$

or in a more compact form:

$$\rho c_p\frac{dT}{dt} = \nabla\cdot(k\nabla T) + q_s \tag{4.B.7b}$$

APPENDIX 4.C MOVING HEAT SOURCE

In deriving Eq. (4.5), let us consider the heat source to be moving along the positive x-axis with a velocity u_x, Figure 4.1. A new set of coordinate axes is attached to the moving heat source. The new

(moving) coordinate system, ξ, y', z', t', is related to the old (fixed) coordinate system, x, y, z, t as follows:

$$\xi = x - u_x t, \quad y' = y, \quad z' = z, \quad t' = t \tag{4.C.1}$$

Then, coordinate transformation gives

$$\frac{\partial T}{\partial x} = \frac{\partial T}{\partial \xi}\frac{\partial \xi}{\partial x} = \frac{\partial T}{\partial \xi}, \quad \frac{\partial^2 T}{\partial x^2} = \frac{\partial^2 T}{\partial \xi^2} \tag{4.C.2a}$$

$$\frac{\partial T}{\partial y} = \frac{\partial T}{\partial y'}\frac{\partial y'}{\partial y} = \frac{\partial T}{\partial y'}, \quad \frac{\partial^2 T}{\partial y^2} = \frac{\partial^2 T}{\partial y'^2} \tag{4.C.2b}$$

$$\frac{\partial T}{\partial z} = \frac{\partial T}{\partial z'}\frac{\partial z'}{\partial z} = \frac{\partial T}{\partial z'}, \quad \frac{\partial^2 T}{\partial z^2} = \frac{\partial^2 T}{\partial'^2} \tag{4.C.2c}$$

Also,

$$\frac{dT}{dt} = \frac{\partial T}{\partial t'}\frac{\partial t'}{\partial t} + \frac{\partial T}{\partial \xi}\frac{d\xi}{dt} = \frac{\partial T}{\partial t'} - u_x \frac{\partial T}{\partial \xi} \tag{4.C.3}$$

Substituting into Eq. (4.3) and noting that $t' = t$ and $\frac{\partial T}{\partial y} = \frac{\partial T}{\partial y'}$ and so on, we get for the moving coordinate system:

$$\frac{\partial T}{\partial t} = \kappa\left(\frac{\partial^2 T}{\partial \xi^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}\right) + u_x \frac{\partial T}{\partial \xi} \tag{4.C.4}$$

APPENDIX 4.D

Thermal and physical properties of some common materials.

Material	Density at 20 °C (kg/m^3)	Melting point (°C)	Heat of fusion (kJ/kg)	Boiling point (°C)	Heat of vaporization (kJ/kg)
Aluminum (Al)	2700	660.4	397	2494	10780
Carbon (C)	—	3540	—	4000	5024
Chromium (Cr)	7190	1875	258-283	2680	6168
Copper (Cu)	8930	1085	205	2595	4729
Gold (Au)	19290	1064	62.8	2857	1698.7
Iron (Fe)	7870	1538	247±7	2870	7018
Lead (Pb)	11340	327.4	22.98–23.38	1750	945.34
Magnesium (Mg)	1738	650	360–377	1107±10	5150–5400
Manganese (Mn)	7300	1244±4	267	2095	4090
Molybdenum (Mo)	10220	2610	270	5560	5123
Nickel (Ni)	8900	1453	293	2730	6196
Niobium (Nb)	8570	2468	290	4927	7490
Silicon (Si)	2330	1410	1807.9	3280	1060
Silver (Ag)	10490	961.9	104.2	2163	2630
Tantalum (Ta)	16600	2996	145–174	5427	4160–4270
Tin (Sn)	7168	231.9	59.5	2770	2400
Titanium (Ti)	4507	1668±10	440	3260	9830
Tungsten (W)	19254	3410	220±36	5700±200	4815
Vanadium (V)	6100	1900±25	—	—	—
Zinc (Zn)	7130	420	100.9	906	1782
Zirconium (Zr)	6530	1852	25	3700	6520

Source: Adapted from Raznjevic (1976), Bardes (1978a, 1978b).

APPENDIX 4.E

Thermal properties of some common materials.[1]

Material	Coefficient of linear expansion at 20–100 °C (μm/m K)	Specific heat at 20 °C (J/kg K)	Thermal conductivity at 20 °C (W/m K)
Aluminum (Al)	23.6	900 at 25 °C	247 at 25 °C
Carbon and graphite (C)	—	708–934	11.6–174
Chromium (Cr)	6.2	459.8	67
Copper (Cu)	16.5	386	398 at 27 °C
Gold (Au)	14.2	128 at 25 °C	317.9 at 0 °C
Iron (Fe)	12 at 100 °C	452	73
Lead (Pb)	29.3	128.7 at 25 °C	34.8
Magnesium (Mg)	25.2	1025	418
Manganese (Mn)	22.8	477 at 25 °C	50.2 at 0 °C
Molybdenum (Mo)	5.2	276	142
Nickel (Ni)	13.3	444	82.9 at 100 °C
Niobium (Nb)	7.31 at 300 °C	270	52.3 at 0 °C
Silicon (Si)	2.8–7.9	678	83.7
Silver (Ag)	19	235 at 25 °C	428
Tantalum (Ta)	6.5	139.1 at 0 °C	54.4
Tin (Sn)	23.1 at 50 °C	205 at 10 °C	62.8 at 0 °C
Titanium (Ti)	10.8 at 100 °C	99.3 at 50 °C	11.4 at −240 °C
Tungsten (W)	4.5 at 100 °C	134	151.2 at 100 °
Vanadium (V)	8.3	498	31 at 100 °C
Zinc (Zn)	15	382	113 at 25 °C
Zirconium (Zr)	5.85	300 at ≈ 100 °C	21.1 at 25 °C

Source: Adapted from Raznjevic (1976); Bardes (1978a, 1978b).

[1] Thermal properties often vary significantly with temperature. The values given are average and/or approximate. For more precise values at a given temperature, the reader should consult one of the references.

APPENDIX 4.F

Reflectance of some common materials.

Material	$\lambda = 0.25\,\mu m$	$\lambda = 1.0\,\mu m$	$\lambda = 10.0\,\mu m$
Aluminum (Al)	86–87	96	97
Chromium (Cr)	67 at $\lambda = 0.3\,\mu m$	63	88 at $\lambda = 4\,\mu m$
Copper (Cu)	25.9	90.1	98.4
Gold (Au)	40	96	96
Iron (Fe)	—	65 at $\lambda = 1.5\,\mu m$	97 at $\lambda = 15\,\mu m$
Lead (Pb)	62 at $\lambda = 0.589\,\mu m$	—	—
Magnesium (Mg)	72 at $\lambda = 0.5\,\mu m$	74	93 at $\lambda = 9\,\mu m$
Molybdenum (Mo)	46 at $\lambda = 0.5\,\mu m$	—	93
Nickel (Ni)	41.3 at $\lambda = 0.3\,\mu m$	—	—
Silicon (Si)	29–39 at $\lambda = 0.13$–$0.2\,\mu m$	—	—
Silver (Ag)	$\approx$30	≈ 95	≈ 95
Tin (Sn)	80 at $\lambda = 0.546\,\mu m$	—	—
Zinc (Zn)	74.7 at $\lambda = 0.5\,\mu m$	53.3	—

Source: Adapted from Bardes (1978a, 1978b).

APPENDIX 4.G

Table of modified Bessel function of the second kind of order zero (Table F-11).

Table F-11. Bessel Functions: $K_0(x)$ and $K_1(x)$ (*Continued*) $K_0(x)$

x	0	1	2	3	4	5	6	7	8	9
0.0	∞	4.721	4.028	3.624	3.337	3.114	2.933	2.780	2.647	2.531
1	2.427	2.333	2.248	2.170	2.097	2.030	1.967	1.909	1.854	1.802
2	1.753	1.706	1.662	1.620	1.580	1.542	1.505	1.470	1.436	1.404
3	1.372	1.342	1.314	1.286	1.259	1.233	1.208	1.183	1.160	1.137
4	1.115	1.093	1.072	1.052	1.032	1.013	0.9943	9761	9584	9412
5	0.9244	9081	8921	8766	8614	8466	8321	8180	8042	7907
6	7775	7646	7520	7397	7277	7159	7043	6930	6820	6711
7	6605	6501	6399	6300	6202	6106	6012	5920	5829	5740
8	5653	5568	5484	5402	5321	5242	5165	5088	5013	4940
9	4867	4796	4727	4658	4591	4524	4459	4396	4333	4271
1.0	4210	4151	4092	4034	3977	3922	3867	3813	3760	3707
1	3656	3605	3556	3507	3459	3411	3365	3319	3273	3229
2	3185	3142	3100	3058	3017	2976	2936	2897	2858	2820
3	2782	2746	2709	2673	2638	2603	2569	2535	2502	2469
4	2437	2405	2373	2342	2312	2282	2252	2223	2194	2166
5	2138	2111	2083	2057	2030	2004	1979	1953	1928	1904
6	1880	1856	1832	1809	1786	1763	1741	1719	1697	1676
7	1655	1634	1614	1593	1573	1554	1534	1515	1496	1478
8	1459	1441	1423	1406	1388	1371	1354	1337	1321	1305
9	1288	1273	1257	1242	1226	1211	1196	1182	1167	1153
2.0	1139	1125	1111	1098	1084	1071	1058	1045	1033	1020
1	1008	*9956	*9836	*9717	*9600	*9484	*9370	*9257	*9145	*9035
2	0.08927	8820	8714	8609	8506	8404	8304	8204	8106	8010
3	7914	7820	7726	7634	7544	7454	7365	7278	7191	7106
4	7022	6939	6856	6775	6695	6616	6538	6461	6384	6309
5	6235	6161	6089	6017	5946	5877	5808	5739	5672	5606
6	5540	5475	5411	5348	5285	5223	5162	5102	5042	4984
7	4926	4868	4811	4755	4700	4645	4592	4538	4485	4433
8	4382	4331	4281	4231	4182	4134	4086	4039	3992	3946
9	3901	3856	3811	3767	3724	3681	3638	3597	3555	3514
3.0	3474	3434	3395	3356	3317	3279	3241	3204	3168	3131
1	3095	3060	3025	2990	2956	2922	2889	2856	2824	2791
2	2759	2728	2697	2666	2636	2606	2576	2547	2518	2489
3	2461	2433	2405	2378	2351	2325	2298	2272	2246	2221
4	2196	2171	2146	2122	2098	2074	2051	2028	2005	1982
5	1960	1938	1916	1894	1873	1852	1831	1810	1790	1770
6	1750	1730	1711	1692	1673	1654	1635	1617	1599	1581
7	1563	1546	1528	1511	1494	1477	1461	1445	1428	1412
8	1397	1381	1366	1350	1335	1320	1306	1291	1277	1262
9	1248	1234	1221	1207	1194	1180	1167	1154	1141	1129
4.0	1116	1104	1091	1079	1067	1055	1044	1032	1021	1009
1	0.009980	9869	9760	9652	9545	9439	9334	9231	9128	9027
2	8927	8829	8731	8634	8539	8444	8351	8259	8167	8077
3	7988	7900	7813	7726	7641	7557	7473	7391	7309	7229
4	7149	7070	6992	6915	6839	6764	6689	6616	6543	6471
5	6400	6329	6260	6191	6123	6056	5989	5923	5858	5794
6	5730	5668	5605	5544	5483	5423	5363	5305	5246	5189
7	5132	5076	5020	4965	4911	4857	4804	4751	4699	4648
8	4597	4547	4497	4448	4399	4351	4304	4257	4210	4164
9	4119	4074	4030	3986	3942	3899	3857	3814	3773	3732
5.0	3691	3651	3611	3572	3533	3494	3456	3419	3382	3345

$$K'_0(x) = -K_1(x)$$

Table F-11. Bessel Functions: $K_0(x)$ and $K_1(x)$ (*Continued*) $K_0(x)$

x	0	1	2	3	4	5	6	7	8	9
5.0	0.003691	3651	3611	3572	3533	3494	3456	3419	3382	3345
1	3308	3272	3237	3202	3167	3132	3098	3065	3031	2998
2	2966	2934	2902	2870	2839	2808	2778	2748	2718	2688
3	2659	2630	2602	2574	2546	2518	2491	2464	2437	2411
4	2385	2359	2333	2308	2283	2258	2234	2210	2186	2162
5	2139	2116	2093	2070	2048	2026	2004	1982	1961	1939
6	1918	1898	1877	1857	1837	1817	1798	1778	1759	1740
7	1721	1703	1684	1666	1648	1630	1613	1595	1578	1561
8	1544	1528	1511	1495	1479	1463	1447	1432	1416	1401
9	1386	1371	1356	1342	1327	1313	1299	1285	1271	1258
6.0	1244	1231	1217	1204	1191	1179	1166	1153	1141	1129
1	1117	1105	1093	1081	1070	1058	1047	1035	1024	1013
2	1002	*9918	*9811	*9706	*9602	*9499	*9398	*9297	*9197	*9099
3	0.0009001	8905	8810	8715	8622	8530	8438	8348	8259	8171
4	8083	7997	7911	7827	7743	7660	7578	7497	7417	7338
5	7259	7182	7105	7029	6954	6880	6806	6734	6662	6591
6	6520	6451	6382	6314	6246	6180	6114	6048	5984	5920
7	5857	5795	5733	5672	5611	5551	5492	5434	5376	5318
8	5262	5206	5150	5095	5041	4987	4934	4882	4830	4778
9	4728	4677	4627	4578	4529	4481	4434	4386	4340	4294
7.0	4248	4203	4158	4114	4070	4027	3984	3942	3900	3858
1	3817	3777	3737	3697	3658	3619	3580	3542	3505	3468
2	3431	3394	3358	3323	3287	3253	3218	3184	3150	3117
3	3084	3051	3019	2987	2955	2924	2893	2862	2832	2802
4	2772	2742	2713	2685	2656	2628	2600	2573	2545	2518
5	2492	2465	2439	2413	2388	2363	2338	2313	2288	2264
6	2240	2216	2193	2170	2147	2124	2102	2079	2057	2036
7	2014	1993	1972	1951	1930	1910	1890	1870	1850	1830
8	1811	1792	1773	1754	1736	1717	1699	1681	1664	1646
9	1629	1611	1594	1578	1561	1545	1528	1512	1496	1480
8.0	1465	1449	1434	1419	1404	1389	1374	1360	1346	1331
1	1317	1303	1290	1276	1263	1249	1236	1223	1210	1198
2	1185	1172	1160	1148	1136	1124	1112	1100	1089	1077
3	1066	1055	1043	1032	1022	1011	10002	*9897	*9793	*9690
4	.00009588	9487	9387	9288	9191	9094	8998	8904	8810	8717
5	8626	8535	8445	8356	8269	8182	8096	8011	7926	7843
6	7761	7679	7598	7519	7439	7361	7284	7208	7132	7057
7	6983	6909	6837	6765	6694	6624	6554	6485	6417	6350
8	6283	6217	6152	6088	6024	5961	5898	5836	5775	5714
9	5654	5595	5536	5478	5420	5364	5307	5252	5197	5142
9.0	5088	5035	4982	4930	4878	4827	4776	4726	4677	4628
1	4579	4531	4484	4437	4390	4344	4299	4254	4209	4165
2	4121	4078	4036	3993	3951	3910	3869	3829	3789	3749
3	3710	3671	3632	3594	3557	3519	3483	3446	3410	3374
4	3339	3304	3270	3235	3202	3168	3135	3102	3070	3038
5	3006	2974	2943	2912	2882	2852	2822	2793	2763	2734
6	2706	2678	2650	2622	2595	2567	2541	2514	2488	2462
7	2436	2411	2385	2360	2336	2311	2287	2263	2240	2216
8	2193	2170	2148	2125	2103	2081	2059	2038	2017	1995
9	1975	1954	1934	1913	1894	1874	1854	1835	1816	1797
10.0	1778	1759	1741	1723	1705	1687	1670	1652	1635	1618

For larger values of the argument, use the auxiliary functions, pages 1057 to 1059.

Source: Korn and Korn (1968)/McGraw-Hill.

PROBLEMS

4.1. The governing equation for conduction heat flow in a solid with a coordinate system fixed at a stationary origin relative to the solid may be expressed as:

$$\frac{dT}{dt} = \kappa \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} \right)$$

Show that for the moving heat source case with a coordinate system moving with the heat source along the x-axis, as shown in Figure 4.1, the corresponding governing equation obtained by a coordinate transformation from the plate to the heat source, with x replaced by ξ, that is, $\xi = x - u_x\, t$, is given by:

$$\frac{\partial T}{\partial t} = \kappa \left(\frac{\partial^2 T}{\partial \xi^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} \right) + u_x \frac{\partial T}{\partial \xi}$$

4.2. Given that

$$T - T_0 = \frac{q}{2\pi kh} \exp\left(-\frac{u_x \xi}{2\kappa} \right) \times K_0 \left(\frac{u_x r}{2\kappa} \right)$$

and that

$$\left(\frac{\partial y}{\partial x} \right)_T = 0$$

and considering temperatures in terms of distance from the fusion zone boundary, show that the peak temperature for a thin plate is given by:

$$\frac{1}{T_\mathrm{p} - T_0} = \frac{\sqrt{2\pi e}\rho c_\mathrm{p} h u_x Y}{q} + \frac{1}{T_\mathrm{m} - T_0}$$

4.3. Given that

$$T - T_0 = \frac{q}{2\pi kr} \exp\left[-\frac{u_x (r + \xi)}{2\kappa} \right]$$

and that

$$\left(\frac{\partial y}{\partial x} \right)_T = 0$$

and considering temperatures in terms of distance from the fusion zone boundary, show that the peak temperature for a thick plate is given by:

$$\frac{1}{T_\mathrm{p} - T_0} = \frac{2\pi k \kappa e}{q u_x} \left[2 + \left(\frac{u_x Y}{2\kappa} \right)^2 \right] + \frac{1}{T_\mathrm{m} - T_0}$$

4.4. The temperature distribution in a semi-infinite plate for a moving point heat source is given as:

$$T - T_0 = \frac{q}{2\pi kr} \exp\left[-\frac{u_x (r + \xi)}{2\kappa} \right]$$

Derive an expression for the cooling rate at points of the plate where the cooling rates are highest. Where do these occur?

4.5. The temperature distribution in a thin plate for a moving line heat source is given as:

$$T - T_0 = \frac{q}{2\pi kh} \exp\left(-\frac{u_x \xi}{2\kappa} \right) \times K_0 \left(\frac{u_x r}{2\kappa} \right)$$

Derive an expression for the cooling rate at points of the plate where the cooling rates are highest.
Hint: For large values of z, $K_0(z)$ can be approximated by:

$$K_0(z) \approx \sqrt{\left(\frac{\pi}{2z}\right)} e^{-z}$$

4.6. For a two-beam source as shown in Figure P4.6 for a thick plate, if each of the heat sources is q, it can be shown that the temperature distribution at any point for this configuration is given by:

$$T - T_0 = \frac{q}{2\pi k}\left[\frac{1}{|\xi|} e^{-\frac{u_x(\xi+|\xi|)}{2\kappa}} + \frac{1}{|\xi'|} e^{-\frac{u_x(\xi'+|\xi'|)}{2\kappa}}\right]$$

And for points behind the second heat source

$$T - T_0 = \frac{q}{2\pi k}\left(\frac{1}{\xi} + \frac{1}{\xi'}\right)$$

$$\xi' = \xi - d$$

$$\xi = x - u_x t$$

On basis of this, show that the cooling rate along the weld centerline, and behind the second heat source is given by:

$$\frac{dT}{dt} = -\left(\frac{2\pi k u_x}{q}(T - T_0)^2 - \frac{q u_x}{\pi k \xi \xi'}\right)$$

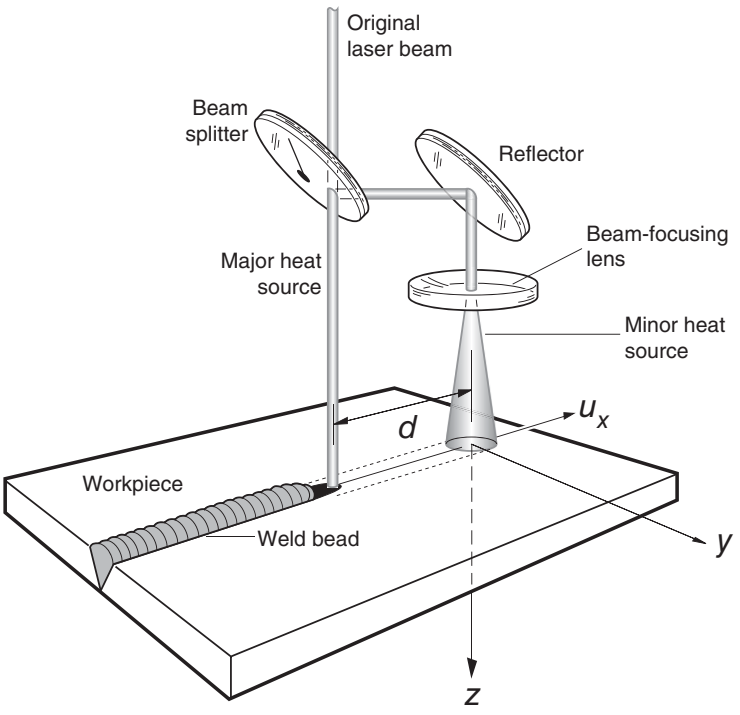


Figure P4.6 Schematic of the multiple-beam laser welding concept.

4.7. Set up the appropriate boundary conditions for the velocity and temperature fields associated with flow in a stationary molten pool with a flat surface.
Hint: Consider the top surface, bottom surface, centerline, side surface, locations far removed from heat source.

4.8. Set up the appropriate boundary conditions for flow in a moving molten pool for:
(a) The top surface of the molten pool.
(b) Along the molten pool center plane.
(c) Far away from the top surface.
(d) Far away from the centerline.

4.9. For the problem in Example 4.2, determine the temperature at a distance 6 mm directly behind the heat source.

4.10. A 5 kW laser beam is used to remelt silicon to produce material for solar cells. The silicon is 100 mm wide and 2 mm thick. The molten pool passes across the width of the silicon, with a velocity of 4 m/min. The ambient temperature is 293 K and the heat transfer efficiency is 50%.
(a) Calculate the maximum cooling rate in the silicon, after cooling to a temperature 400 °C below the melting temperature. At what location is this cooling rate achieved?
(b) At what distance from the fusion zone boundary is a peak temperature of 1580 K achieved?

4.11. A manufacturing engineer is responsible for reducing the cooling rates that occur during laser cladding of a high hardenability material. If the workpiece is determined to be relatively thin,
(a) Identify three primary ways that he or she can reduce the cooling rates.
(b) Which of these three approaches would be most effective? Explain.

4.12. Two wide and thin pieces of cold worked magnesium plates, 2.5 mm thickness, are to be butt-welded using a 5 kW laser.
(a) Calculate the surface temperature at a point 6 mm behind the laser beam and 3 mm to one side.
(b) Estimate the size of the HAZ for this process, assuming recrystallization occurs at a temperature equal to half the melting temperature.
You may assume that a centerline cooling rate (at 500 °C) has been determined to be 3500 °C/s, the initial workpiece temperature is 25 °C, and the process has a heat transfer efficiency of 75%.
(c) Now assume that the workpieces have been preheated to a temperature of 75 °C. Recalculate the surface temperature at the same point (assume that the centerline cooling rate remains unchanged). Does it change? Also, does this result in a larger or smaller HAZ?

4.13. An airbag assembly has a nylon strip attached to a 2 mm thin stainless steel plate which is to be welded to another steel component. The nylon strip is 3 cm away from the weld fusion boundary and parallel to it. The melting point of nylon is 260 °C, and welding is done using a 5 kW CO_2 laser. If the heat transfer efficiency is 80% and the ambient temperature is 25 °C, determine
(a) The welding speed that will not result in melting of the nylon.
(b) The range of plate thicknesses that can be considered thin.

BIBLIOGRAPHY

Abramowitz M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions*, National Bureau of Standards Applied Mathematics Series 55, Washington, D.C.

Adams, C.M. (1958). Cooling rates and peak temperatures in fusion welding. *Welding Journal* 37: 210s–215s.

Anisimov, S.I., Kapeliovich, B.L., and Perel'man, T.L. (1974). Electron emission from metal surfaces exposed to ultrashort laser pulses. *Soviet Physics JETP* 39: 375–377.

Arpaci, V.S. (1966). *Conduction Heat Transfer*. Reading, Massachusetts: Addison Wesley Publishing Co.

Bardes, B.P. (ed.) (1978a). *Metals Handbook, Properties and Selection: Irons and Steels*, 9the, vol. 1. Ohio: American Society for Metals.

Bardes, B.P. (ed.) (1978b). *Metals Handbook, Properties and Selection: Nonferrous Alloys and Pure Metals*, 9the, vol. 2. Ohio: American Society for Metals.

Carslaw, H.S. and Jaeger, J.C. (1959). *Conduction of Heat in Solids*. Oxford: Clarendon Press.

Chan, C., Mazumder, J., and Chen, M.M. (1984). A two-dimensional transient model for convection in laser melted pool. *Metallurgical Transactions A* 15A: 2175–2184.

Chande, T. and Mazumder, J. (1984). Estimating effects of processing conditions and variable properties upon pool shape, cooling rates, and absorption coefficient in laser welding. *Journal of Applied Physics* 56: 1981–1986.

Chen, J.K., Beraun, J.E., and Tzou, D.Y. (2002). Thermomechanical response of metals heated by ultrashort-pulsed lasers. *Journal of Thermal Stresses* 25: 539–558.

Chen, J.K., Latham, W.P., and Beraun, J.E. (2005). The role of electron-phonon coupling in ultrafast laser heating. *Journal of Laser Applications* 17 (1): 63–68.

Cline, H.E. and Anthony, T.R. (1977). Heat treating and melting material with a scanning laser or electron beam. *Journal of Applied Physics* 48: 3895–3900.

Heiple, C.R. and Roper, J.R. (1981). Effect of selenium on gtaw fusion zone geometry. *Welding Journal* 60: 143s–145s.

Heiple, C.R., Roper, J.R., Stagner, R.T., and Aden, R.J. (1983). Surface active element effects on the shape of GTA, laser, and electron beam welds. *Welding Journal* 62: 72s–77s.

Kannatey-Asibu, E. Jr. (1991). Thermal aspects of the split-beam laser welding concept. *ASME Journal of Engineering Materials and Technology* 113: 215–221.

Kim, Y.-S., McEligot, D.M., and Eagar, T.W. (1991). Analyses of electrode heat transfer in gas metal arc welding. *Welding Journal* 70: 20s–31s.

Korn, G.A. and Korn, T.M. (1968). *Mathematical Handbook for Scientists and Engineers*. New York: McGraw-Hill.

Korte, F., Serbin, J., Koch, J. et al. (2003). Towards nanostructuring with femtosecond laser pulses. *Journal of Applied Physics A* 77: 229–235.

Kou, S. and Wang, Y.H. (1986). Weld pool convection and its effect. *Welding Journal* 65: 63s–70s.

Kou, S. (2003). *Welding Metallurgy*, 2nde. Hoboken, New Jersey: Wiley.

Lancaster, J.F. (1965). *Metallurgy of Welding, Brazing, and Soldering*. London: George Allen & Unwin Ltd.

Liu, Y.N. and Kannatey-Asibu, E. Jr. (1993). Laser beam welding with simultaneous Gaussian laser preheating. *ASME Journal of Heat Transfer* 115: 34–41.

Mazumder, J. and Steen, W.M. (1980). Heat transfer model for cw material processing. *Journal of Applied Physics* 51 (2): 941–947.

Mazumder, J. (1987). An overview of melt dynamics in laser processing. *High Power Lasers* 801: 228–241.

Oreper, G.M. and Szekeley, J. (1984). Heat and fluid-flow phenomena in weld pools. *Journal of Fluid Mechanics* 147: 53–79.

Paul, A. and Debroy, T. (1988). Free surface flow and heat transfer in conduction mode laser welding. *Metallurgical Transactions B* 19B: 851–858.

Qiu, T.Q. and Tien, C.L. (1992). Short-pulse laser heating on metals. *International Journal of Heat and Mass Transfer* 35 (3): 719–726.

Qiu, T.Q. and Tien, C.L. (1993). Heat transfer mechanisms during short-pulse laser heating of metals. *ASME Journal of Heat Transfer* 115: 835–841.

Ramanan, N. and Korpela, S.A. (1990). Fluid dynamics of a stationary weld pool. *Metallurgical Transactions A* 21A: 45–56.

Rapkin, D.M. (1959). Temperature distribution through the weld pool in the automatic welding of aluminum. *British Welding Journal* 6: 132–137.

Raznjevic, K. (1976). *Handbook of Thermodynamic Tables and Charts*. New York: McGraw-Hill Book Company.

Rosenthal, D. (1941). Mathematical theory of heat distribution during welding and cutting. *Welding Journal* 20: 220s–234s.

Rosenthal, D. (1946). The theory of moving sources of heat and its application to metal treatments. *Trans ASME* 68: 849–866.

Russo, A.J., Akau, R.L., and Jellison, J.L. (1990). Thermocapillary flow in pulsed laser beam weld pools. *Welding Journal* 69: 23s–29s.

Sherman, F.S. (1990). *Viscous Flow*. New York, NY: McGraw-Hill Publishing Co.

Sanders, D.J. (1984). Temperature distributions produced by scanning gaussian laser beams. *Applied Optics* 23: 30–35.

Walsh, D.W. and Savage, W.F. (1985). Technical note: autogenous gta weldments - bead geometry variations due to minor elements. *Welding Journal* 68: 59s–62s.

Wells, A.A. (1952). Heat flow in welding. *Welding Journal* 31: 263s–267s.

Wylie, C.R. and Barrett, L.C. (1982). *Advanced Engineering Mathematics*, 5th ed. New York: McGraw-Hill, Inc.

Yilbas, B.S. (2004). Convergence of electron kinetic, two-temperature, and one-temperature models for laser short-pulse heating. *Journal of Applied Physics A* 79: 1775–1782.

Zacharia, T., Eraslan, A.H., and Aidun, D.K. (1988). Modeling of autogenous welding. *Welding Journal* 68: 53s–62s.

Zacharia, T., David, S.A., Vitek, J.M., and Debroy, T. (1989). Weld pool development during GTA and laser beam welding of type 304 stainless steel, parts 1 and 2. *Welding Journal* 69, 499s–509s, 510s–519s.

5 The Microstructure

The heat flow associated with laser processing does affect the microstructure of the part that is produced. In this chapter, basic microstructural concepts and discontinuity (defect) formation, as they relate to laser processing, are discussed. Defects are discussed in this chapter, since they are closely tied to the microstructural changes that occur during processing.

The chapter starts with a brief outline of the major microstructural zones that occur in the part during processing, addressing the fusion zone, zone of partial melting, and the heat-affected zone (HAZ). Under fusion zone, basic concepts of solidification are introduced, starting with nucleation, where the conditions necessary for the formation of a stable nucleus are outlined. The concepts of homogeneous and heterogeneous nucleation are presented. This is followed by a discussion on the growth of the stable nuclei to form crystals or grains.

The microstructure that results as the molten material solidifies is presented next, outlining the principal types of structure that may result, that is, planar, cellular, or dendritic, and the criterion that determines which structure is most likely to form. Techniques for controlling nucleation and also for grain refinement are then presented.

The zone of partial melting is then discussed, followed by the HAZ that results for pure metals, precipitation-hardening and non-ferrous alloys, and steels. Next, common discontinuities that can result during laser processing are discussed. These include porosity, cracking, incomplete penetration, lack of fusion, and undercutting.

Analyses in this chapter are based on fundamental concepts that lead to closed-form solutions for first-order approximations. Several methods are available for more precise microstructural analyses using numerical simulation. Among these are cellular automata, Monte Carlo, and phase field methods.

The process microstructure is discussed next.

5.1 PROCESS MICROSTRUCTURE

The processes that involve fusion (laser cutting, welding, surface melting, and cladding) have four distinct zones in the region where the heat is applied (Figure 5.1). These are as follows:

1. Fusion zone.
2. Zone of partial melting.
3. HAZ.
4. Base material.

Principles of Laser Materials Processing: Developments and Applications, Second Edition. Elijah Kannatey-Asibu, Jr.
© 2023 John Wiley & Sons, Inc. Published 2023 by John Wiley & Sons, Inc.
Companion Website: www.wiley.com/go/Kannatey-Asibu/PrinicplesLaserMaterialsProcessing

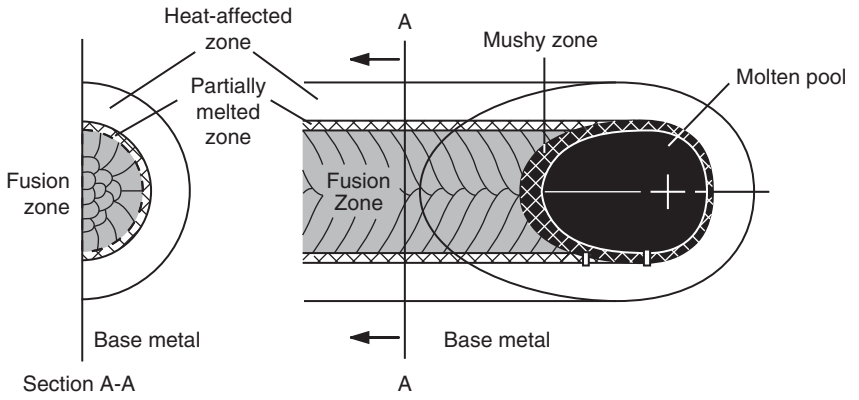


Figure 5.1 Major microstructural zones associated with laser processing. Source: From Kou (1987)/with permission of John Wiley & Sons, Inc.

The first three zones will be discussed individually in the following sections. For processes that do not involve fusion (laser heat treatment and laser forming), microstructural changes occur only in the HAZ.

5.1.1 Fusion Zone

The fusion zone is the portion of the material which melts. The discussion on the fusion zone starts with an introduction to the solidification process, focusing on grain nucleation and growth. This is followed by the microstructure of the fusion zone, nucleation, and grain refinement in the molten pool. Solidification is discussed in greater detail in the next chapter.

5.1.1.1 Initial Solidification Solidification generally occurs in two stages:

1. Nucleation, which occurs when a small piece of stable solid (nucleus) forms from the liquid.
2. Growth, which occurs as atoms from the liquid are attached to the nucleus.

Nucleation A nucleus is the smallest particle of the new phase (solid in this case) that can exist in a stable form. Nucleation will generally occur when the part is cooled below an equilibrium temperature, T_e. This is the temperature at which the free energies of the two phases (solid and liquid) are equal. Below it, the solid is the more stable phase, since its free energy is then lower than that of the liquid phase.

The process of nucleation can occur in two forms:

1. Homogeneous nucleation.
2. Heterogeneous nucleation.

In homogeneous nucleation, the nuclei of the second-phase form within the bulk or matrix of the original phase, say the liquid, in the case of solidification. This rarely occurs in practice, and for it to be possible, there must be extensive supercooling. In heterogeneous nucleation, the nuclei either form on foreign particles suspended in the liquid or on the walls of the container, and the amount of undercooling necessary for it to occur is much smaller than that required for homogeneous nucleation. In either case, each nucleus constitutes the basis for a crystal or grain, which may subsequently grow.

For homogeneous nucleation, the total free energy change, ΔG, associated with the formation of a spherical solid nucleus in a liquid medium consists of two components:

1. The bulk free energy change.
2. The surface energy change.

Thus, the total free energy change is given by:

$$\Delta G = \frac{4}{3}\pi r^3 \Delta G_v + 4\pi r^2 \gamma_{nl} \tag{5.1}$$

where r is the radius of the nucleus, ΔG_v denotes the bulk free energy change per unit volume of transformed material, and γ_{nl} stands for surface energy per unit area between the solidified nucleus and liquid (molten metal).

Typical surface energy values for some common metals are shown in Table 5.1.

ΔG_v may be expressed as:

$$\Delta G_v = \frac{L_{mv}\Delta T_e}{T_e} \tag{5.2}$$

where L_{mv} is the latent heat of fusion per unit volume, T_e = equilibrium temperature = T_m, $\Delta T_e = T_e - T$ = degree of supercooling, and T is the temperature of the molten metal.

Since heat is evolved during solidification, L_{mv} is negative, and thus the bulk free energy change ΔG_v will be negative at all temperatures below the solidification temperature, that is, $T < T_e$, while the surface energy change is always positive.

The variation of the two terms in Eq. (5.1) with the radius of the nucleus is shown in Figure 5.2. It is evident from Figure 5.2 that there is a free energy barrier to the nucleation process. In other words, a critical nucleus size, r_c, is necessary for the process to continue. At the critical radius, the system is in metastable equilibrium. A small decrease in the radius at that point will result in the surface energy term dominating, and the nucleus will redissolve, since the total free energy decreases with decreasing radius in the region of the curve below the critical radius. An increase in the radius at the metastable state will result in a continued decrease in free energy with increasing radius as the bulk free energy term dominates. The continued decrease in free energy is a necessary condition for the formation of a stable nucleus. A stable nucleus is then formed.

To determine the critical radius, r_c, at which a nucleus becomes stable, Eq. (5.1) is differentiated with respect to r and equated to zero, giving

$$r_c = -\frac{2\gamma_{nl}}{\Delta G_v} = -\frac{2\gamma_{nl}T_e}{L_{mv}\Delta T_e} \tag{5.3}$$

Table 5.1 Surface energies, γ_{nl}, of some common metals.

Metal	γ_{nl} (J/m^2)
Copper (Cu)	0.177
Gold (Au)	0.132
Iron (Fe)	0.204
Lead (Pb)	0.033
Manganese (Mn)	0.206
Nickel (Ni)	0.255
Silver (Ag)	0.126
Tin (Sn)	0.0545

Source: From Hosford (2005)/Taylor & Francis.

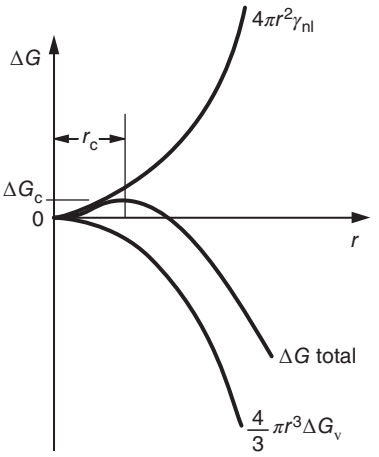


Figure 5.2 Free energy change associated with the formation of a spherical nucleus.

and the corresponding critical free energy change, ΔG_c,

$$\Delta G_c = \frac{16\pi\gamma_{nl}^3}{3(\Delta G_v)^2} = \frac{16\pi\gamma_{nl}^3}{3L_{mv}^2}\frac{T_e{}^2}{\Delta T_e{}^2} \tag{5.4}$$

Example 5.1 Two pieces of silver are butt-welded together using a laser beam, without any filler metal (autogenous welding).

(a) Determine the critical radius necessary for the formation of a stable nucleus if the weld pool was supercooled to a temperature of 724 °C.
(b) Determine the critical free energy change necessary for forming the stable nucleus.

Solution:
From Appendix 4D, the properties of silver are approximated as:
Average density $\rho = 10{,}490\,\text{kg/m}^3$
Heat of fusion $L_m = 104.2\,\text{kJ/kg}$
Melting temperature $T_m = 962\,^\circ\text{C} = 1235\,\text{K}$.

(a) The critical radius is given by Eq. (5.3):

$$\begin{aligned} r_c &= -\frac{2\gamma_{nl}}{\Delta G_v} = -\frac{2\gamma_{nl}T_e}{L_{mv}\Delta T_e} \\ &= -\frac{2\times 0.126\times 1235}{(-104.2\times 10^3)\times 10490\times(962-724)} \\ &= 1.196\times 10^{-3}\ \mu\text{m} = 1.196\ \text{nm} \end{aligned}$$

(b) The critical free energy change is given by:

$$\Delta G_c = \frac{16\pi\gamma_{nl}^3}{3L_{mv}^2}\frac{T_e^2}{\Delta T_e^2}$$

$$= \frac{16\pi \times 0.126^3}{3 \times (104.2 \times 10^3 \times 10490)^2} \frac{1235^2}{(962 - 724)^2}$$
$$= 7.6 \times 10^{-19}\, J.$$

In heterogeneous nucleation during cladding for example, since the nucleation would be on the surface of the unmelted base metal, any nucleus formed would not be completely spherical in shape, but rather can be approximated as a spherical cap (Figure 5.3). For such a shape, the total free energy change associated with the nucleation would be given by:

$$\Delta G = \left[\frac{4}{3}\pi r^3 \Delta G_v + 4\pi r^2 \gamma_{nl}\right] f(\varphi) \tag{5.5}$$

where ϕ is the wetting angle (Figure 5.3), and

$$f(\varphi) = \frac{1}{4}(2 + \cos\varphi)(1 - \cos\varphi)^2 \tag{5.6}$$

The contact or wetting angle ϕ is obtained by considering the respective surface energies between the fusion zone nucleus and solid base metal, γ_{ns}; the solid base metal and liquid (molten) pool, γ_{sl}; and the fusion zone nucleus and liquid (molten) pool, γ_{nl}. For the spherical cap nucleus shown in Figure 5.3, equilibrium of horizontal forces acting on the nucleus gives

$$\gamma_{sl} = \gamma_{ns} + \gamma_{nl}\cos\varphi \quad \text{or} \quad \cos\varphi = \frac{\gamma_{sl} - \gamma_{ns}}{\gamma_{nl}} \tag{5.7}$$

The corresponding critical free energy change, ΔG_c, for this case is given by:

$$\Delta G_c = \frac{16\pi\gamma_{nl}^3}{3(\Delta G_v)^2} f(\varphi) = \frac{16\pi\gamma_{nl}^3}{3L_{mv}^2} \frac{T_e^2}{\Delta T_e^2} f(\varphi) \tag{5.8}$$

The number of nuclei per unit volume, n_v, that are formed at any temperature, T, may be expressed as:

$$n_v = se^{\left(-\frac{\Delta G_c}{k_B T}\right)} \tag{5.9}$$

where k_B is Boltzmann's constant and s is the number of sites available for nuclei formation.

The rate of nucleation $\dot{n}$ depends on two principal parameters:

1. The number of nuclei with sufficient energy to be stable.
2. The ability of atoms to diffuse to the potentially stable nucleus.

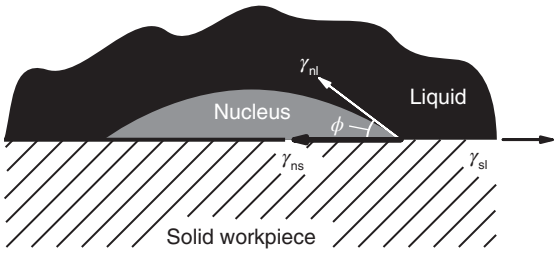


Figure 5.3 Heterogeneous nucleation at the base metal–molten pool interface.

It is normally expressed as:

$$\dot{n} = \frac{k_B T}{2\pi h_p V_m} e^{\left(-\frac{\Delta G_c}{k_B T}\right)} e^{\left(-\frac{\Delta E_d}{k_B T}\right)} = \frac{k_B T}{2\pi h_p V_m} e^{\left(-\frac{[\Delta G_c + \Delta E_d]}{k_B T}\right)} \tag{5.10}$$

where h_p is Planck's constant, V_m is the volume of a molecule, and ΔE_d is the activation energy for diffusion.

The dependence of both the number of nuclei and nucleation rate on temperature is such that the rate is low just below the equilibrium temperature, increases to a maximum as the degree of supercooling increases, and then decreases to zero toward absolute zero. This is because the diffusion rate is relatively fast at temperatures just below T_e. However, since the supercooling, ΔT_e, is then small, the critical free energy change necessary for nucleation is then large (Eq. (5.4)). Thus, only a few nuclei attain sufficient energy to reach the critical size. At relatively lower temperatures where ΔT_e is large, the critical free energy change for nucleation is correspondingly small. However, the diffusion rate is then low, and atoms can no longer move about easily to form nuclei. The nucleation rate is then low. A maximum occurs in the nucleation rate at intermediate temperatures where the diffusion rate is relatively high, and the critical free energy change for nucleation is not too large.

The nucleation principles outlined in the preceding paragraphs are also applicable to solid–solid transformations, for example, austenite to pearlite transformation, which may occur in steels after solidification is completed and the fusion zone continues to cool. However, the resulting nucleation rates are generally much smaller in this case for a number of reasons:

1. Much lower diffusion rates in solids.
2. Higher interfacial energies in the solid state, resulting in higher critical free energy change for the formation of a nucleus.
3. Lower bulk free energy change in the solid, which also increases the critical free energy change.
4. Since the solid does not flow very easily to accommodate volume changes associated with the formation of a nucleus, there is strain energy associated with the formation of the nucleus. That increases the total free energy change necessary for the nucleus formation, since the strain energy term is positive. This term is absent when there is liquid present since the liquid flows easily to accommodate any deformation or volume change associated with the transformation.

Example 5.2 For the conditions in Example 5.1,

(a) Determine the number of stable nuclei per unit volume. Assume the density of nucleation sites is 2.5×10^{30} per m^3.

(b) Determine the nucleation rate. Assume the radius of a silver atom is 0.144 nm, and the activation energy for diffusion in silver is 44.1 kcal/mol = 184.5 kJ/mol (Shewmon, 1983, p. 66).

Solution:

(a) The number of stable nuclei per unit volume is given by:

$$n_v = s e^{\left(-\frac{\Delta G_c}{k_B T}\right)}$$

Using the results of Example 5.1, we have

$$n_v = 2.5 \times 10^{30} \times e^{\left(-\frac{7.6 \times 10^{-19}}{1.38 \times 10^{-23} \times (724 + 273)}\right)}$$

$$= 2.56 \times 10^6 \text{ per } m^3$$

(b) The volume of the silver atom is

$$V_{\mathrm{m}} = \frac{4}{3}\pi r^3 = \frac{4}{3}\pi(0.144 \times 10^{-9})^3 = 12.61 \times 10^{-30}\ \mathrm{m}^3$$

Since the activation energy is given in units of energy per mole, the nucleation rate per unit volume can be written in terms of the gas constant as:

$$\dot{n} = \frac{k_{\mathrm{B}}T}{2\pi h_{\mathrm{p}} V_{\mathrm{m}}} e^{\left(-\frac{\Delta G_{\mathrm{c}}}{k_{\mathrm{B}}T}\right)} e^{\left(-\frac{\Delta E_{\mathrm{d}}}{R_{\mathrm{g}}T}\right)}$$

where $R_{\mathrm{g}} = 1.98$ cal/mol is the gas constant. The nucleation rate then becomes

$$\begin{aligned}\dot{n} &= \frac{1.38 \times 10^{-23} \times 997}{2\pi \times 6.63 \times 10^{-34} \times 12.61 \times 10^{-30}} e^{\left(-\frac{7.6\times10^{-19}}{1.38\times10^{-23}\times997}\right)} e^{\left(-\frac{44.1\times10^{3}}{1.98\times997}\right)} \\ &= 5.32 \times 10^{7}\ \text{per}\ (\mathrm{m}^3\mathrm{s})\end{aligned}$$

Growth The growth of a stable nucleus involves movement of atoms to and attachment to the nucleus. The growth process is thus diffusion-controlled.

Since the base metal is at a lower temperature than the molten pool during laser processing, solidification of the molten metal would ordinarily begin with the formation of solid nuclei at the unmelted surface of the base metal in contact with the molten pool. The nuclei that are formed would then grow into the molten metal, parallel to the direction of heat flow from the molten metal. Such is the case in casting, for example, where nucleation occurs on the mold walls at a temperature just below the melting temperature. On the other hand, initial solidification of a molten pool, in welding or surface melting, for example, is usually *epitaxial*. In this case, atoms from the molten metal are deposited directly on the existing crystals of the unmelted base metal, thereby extending the substrate crystalline structure without changing it (Figure 5.4). This is especially true in autogenous welding and surface melting. For such processes, since the initial grains in the fusion zone grow epitaxially from the substrate grains, the initial crystal structure of the fusion zone is inherited directly from that of the base metal. In other words, the structure and crystallographic orientation of the base metal grains at the fusion zone interface continue into the fusion zone. Thus, the nucleation stage is bypassed at the interface.

To understand the driving mechanism for epitaxial growth, the basic kinetics for phase transformation will be considered. The molten metal in autogenous welding or surface melting is normally

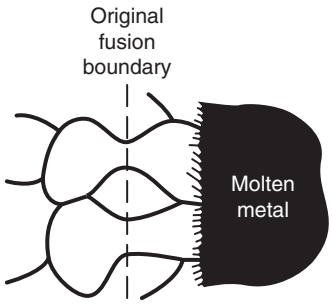


Figure 5.4 Epitaxial solidification from the base metal grains at the fusion line.

of the same material as the substrate and is also in intimate contact with it. Thus, it completely wets the surface of the base metal, resulting in a zero value for ϕ in Eq. (5.6). It is thus evident that $f(\phi)$ is also zero, and consequently, so is ΔG_c. In other words, the critical free energy change required for the formation of a stable nucleus of solid material from the molten metal is zero. Grain growth thus occurs spontaneously from the base metal since there is no resistance to its initiation, that is, the original grains in the base metal in contact with the molten pool continue to grow epitaxially into the pool, without nucleation.

5.1.1.2 Microstructure In this section, we discuss the various microstructures that result from solidification

Structure of the Solidified Metal A pure metal solidifies at a constant temperature, with the interface between the solid and liquid media being planar (Figures 5.5a and 5.6a). For an alloy, the temperature

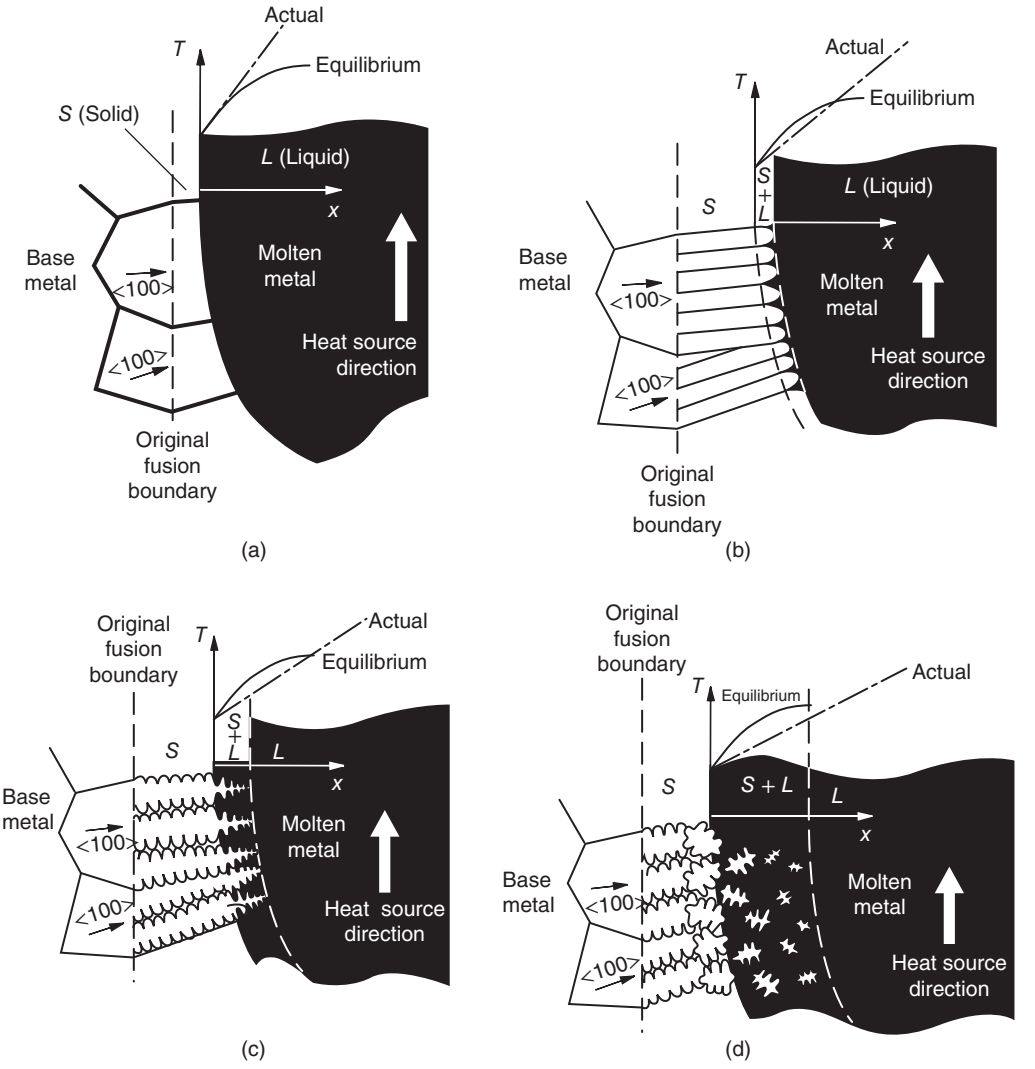


Figure 5.5 (a) Planar, (b) cellular, (c) columnar dendritic, and (d) equiaxed dendritic solid–liquid interfaces during solidification. Source: From Kou (1987)/with permission of John Wiley & Sons, Inc.

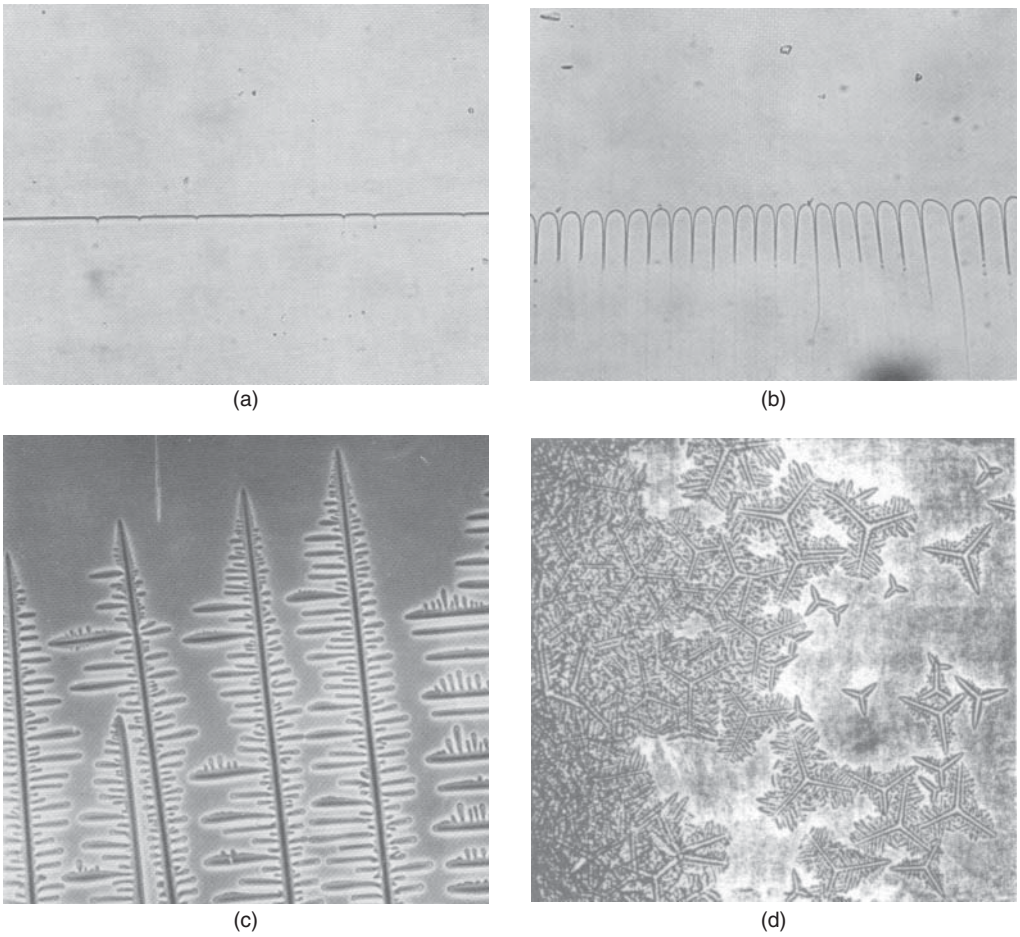


Figure 5.6 Pictorial illustrations of (a) planar, (b) cellular, (c) columnar dendritic, and (d) equiaxed dendritic growth. Source: Kou (1987)/with permission of John Wiley & Sons, Inc.

may decrease as solidification proceeds, unless a eutectic is formed. The interface in this case can thus be planar, cellular, or dendritic, with the dendritic structure being further classified as columnar or equiaxed (Figures 5.5a–d and 5.6a–d). Which one prevails is influenced by the solute content of the molten metal and solidification conditions.

In the initial stages of the solidification of alloys, the low-melting elements are rejected by the freezing solid. Thus, the enriched liquid directly ahead of the solid–liquid interface has a lower freezing point than the previously solidified material. Since the low-melting atoms cannot diffuse rapidly away from the interface, they tend to retard growth. However the liquid temperature is higher, farther away from the interface, that is, toward the center of the molten metal. Thus, portions of the interface protrude or grow rapidly into the liquid, resulting in columns that grow from the initially solidified material, forming a *cellular* or *columnar* structure. For even lower temperature gradients at the interface, branches can grow from the main columns, resulting in a *dendritic* structure. An individual grain may contain a number of dendrites.

The planar interface in an alloy breaks down to form either a cellular or dendritic structure if the following condition holds

$$\frac{G_{\mathrm{Tl}}}{SR} \leq \frac{\Delta T_{\mathrm{f}}}{D_{\mathrm{l}}} \tag{5.11}$$

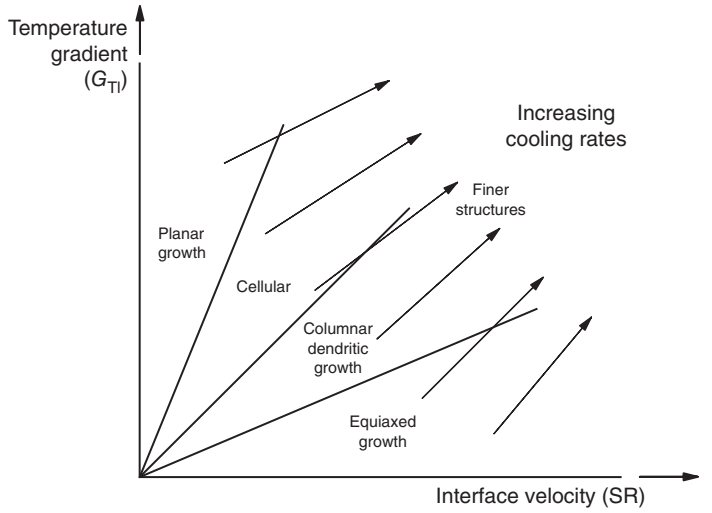


Figure 5.7 Dependence of solidification structure on liquid temperature gradient and solidification rate.

where D_l is diffusion coefficient of the solute in the liquid, G_{Tl} is temperature gradient of the liquid near the solid–liquid interface, *SR* stands for solidification rate, $\Delta T_f = T_l - T_s$ = equilibrium freezing range, T_l is liquidus temperature, and T_s denotes solidus temperature.

The tendency toward a dendritic structure increases with a decreasing value of the ratio on the left-hand side of the equation, that is, as the solidification rate increases, or the temperature gradient of the liquid near the solid–liquid interface decreases. Further reductions in the ratio again result in an increased tendency from a columnar to an equiaxed dendritic structure. This is illustrated schematically in Figure 5.7. The structure that is determined by the solid–liquid interface exists within the individual grains and is commonly referred to as the substructure or subgrain structure. The formation of a dendritic structure is illustrated schematically in Figure 5.8.

The product of the average temperature gradient, G_{Tl}, and interface velocity or solidification rate, *SR*, is a measure of the average cooling rate, *CR*:

$$CR = G_{Tl} \cdot SR \approx \frac{T_l - T_s}{t_s} \tag{5.12}$$

where t_s is local solidification time.

Structural Variation Within Fusion Zone The microstructure or mode of solidification of the fusion zone varies as one moves from the fusion boundary toward the centerline. This results primarily from variations in the solidification rate and temperature gradient across the fusion zone. To further understand this, consider the relationship between the solidification rate **SR** at any point along the molten pool boundary and traverse velocity vector $\mathbf{u}_x$, which, as can be inferred from Figure 5.9, is given by:

$$\mathbf{SR} = \mathbf{u}_x \cos\psi \tag{5.13}$$

where ψ is the angle between the traverse velocity vector and the direction of solidification at any point along the molten pool boundary.

Thus, the solidification rate *SR* is lowest (almost zero) along the fusion line where $\psi \approx 90°$ and maximum along the centerline where $\psi \approx 0°$. Also, the elongated nature of the molten pool is such that the temperature gradient G_{Tl} is highest along the fusion line and lowest along the centerline.

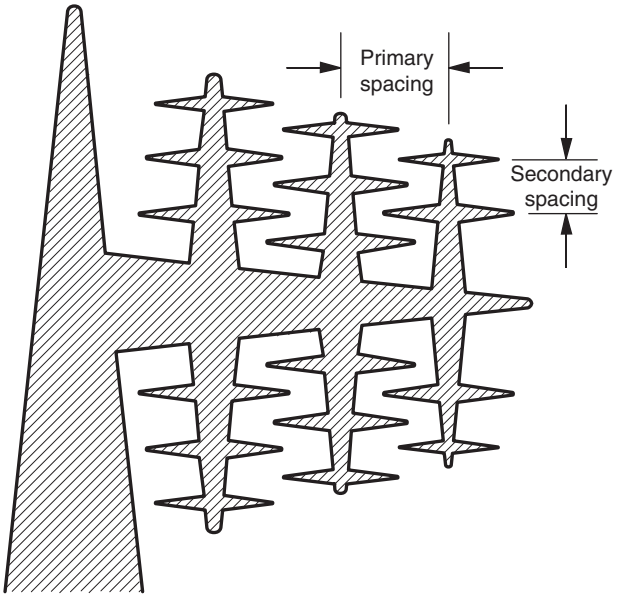


Figure 5.8 Schematic of dendrite formation.

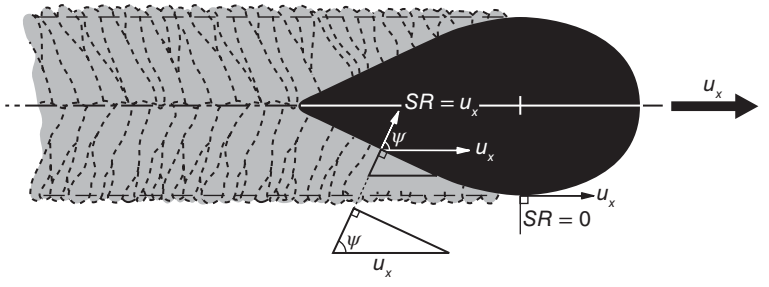


Figure 5.9 Variation of solidification rate along the molten pool edge. Source: From Kou (1987)/with permission of John Wiley & Sons, Inc.

Thus, it is evident from Eq. (5.11) that initial epitaxial growth of the grains along the fusion line is more likely to have a planar front (Figures 5.5a and 5.6a). As one moves inward along the molten pool boundary, the temperature gradient decreases, while the solidification rate increases, with the result that there is an increased tendency toward cellular growth, resulting in columnar grains (Figures 5.5b and 5.6b). Further reductions in the temperature gradient and increases in the solidification rate toward the centerline may result in subsequent transition to a dendritic structure in the fusion zone (Figures 5.5c and 5.6c). This is especially so along the centerline where the temperature gradient is lowest and where the solidification rate is highest. An equiaxed dendritic structure (Figures 5.5d and 5.6d) may thus develop at very high traverse velocities and normally occurs in the final stages of solidification, for example, in a weld crater.

Grain Orientation The direction of grain growth during solidification is controlled by two factors:

1. The maximum temperature gradient direction.
2. The preferred crystallographic growth direction, which is the ⟨100⟩ direction for face-centered and body-centered cubic metals and the ⟨1010⟩ direction for hexagonal close-packed metals.

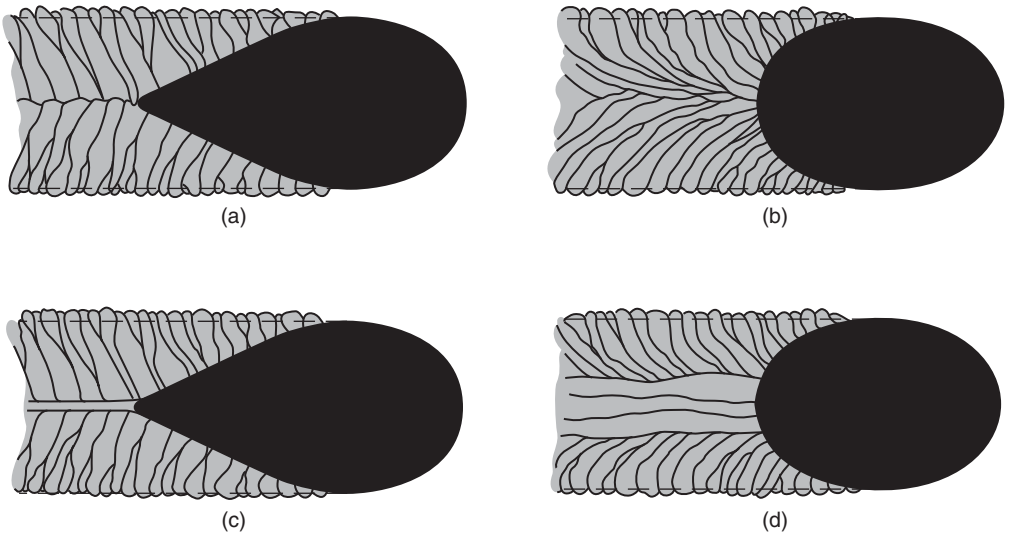


Figure 5.10 (a) Grain structure at high traverse velocities. (b) Grain structure at low traverse velocities. (c) Banded grain structure at high traverse velocities. (d) Banded grain structure at low traverse velocities. Source: From Kou (1987)/with permission of John Wiley & Sons, Inc.

Considering the temperature gradient effect, the grains have a tendency to grow in the direction of steepest temperature gradient. Since, in the case of solidification, this direction is normal to the molten pool boundary, the general direction of grain orientation is also normal to the molten pool boundary. Furthermore, since the shape of the molten pool depends on the traverse speed, so does the direction of grain orientation. Thus, at high traverse velocities, the columnar grains grow straight but are inclined to the fusion boundary since they tend to grow in a direction perpendicular to the elongated molten pool boundary as shown in Figure 5.10a. However, at low traverse velocities, they tend to curve toward the middle of the molten pool, to become perpendicular to the elliptical pool boundary (Figure 5.10b). In situations where other grains are nucleated in the middle of the molten pool, these grow in the direction of motion. This is because the pool boundary toward the center is normal to the direction of motion at any speed. Thus, for such cases, the elongated shape at high speeds results in a narrow band of grains in the middle (Figure 5.10c), while the elliptical shape at lower speeds results in a wider band (Figures 5.10d and 5.11).

In considering the preferred crystallographic growth direction, it is found that in the initial stages of epitaxial growth, some of the original base metal crystals may have their preferred growth direction in the same direction as the direction of steepest temperature gradient. Such grains will have a greater driving force for growth and are thus more likely to grow at the expense of those which have a less favorable orientation (Figure 5.12). Furthermore, crystals sometimes change orientation during growth to enable the two directions (preferred growth and steepest gradient) to be coincident and thus facilitate growth.

5.1.1.3 Nucleation and Grain Refinement in Molten Pool When new grains are nucleated in the molten pool, they tend to form an equiaxed rather than a columnar grain structure. The new grains

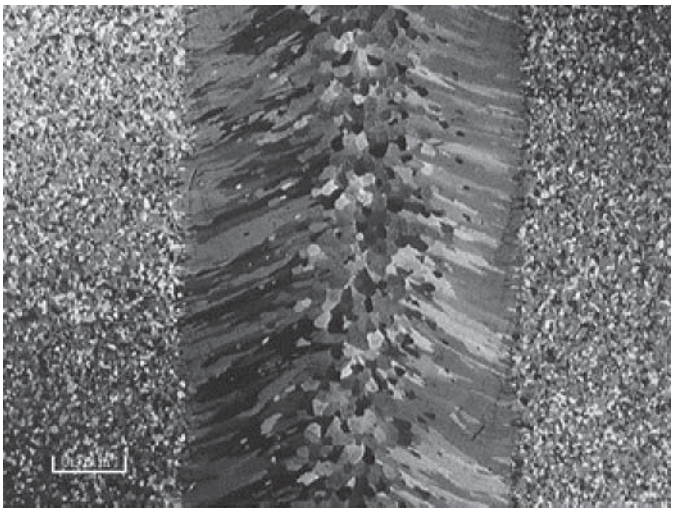


Figure 5.11 Pictorial illustration of a banded grain structure. Source: Friedman et al. (2000), with permission of Springer Nature.

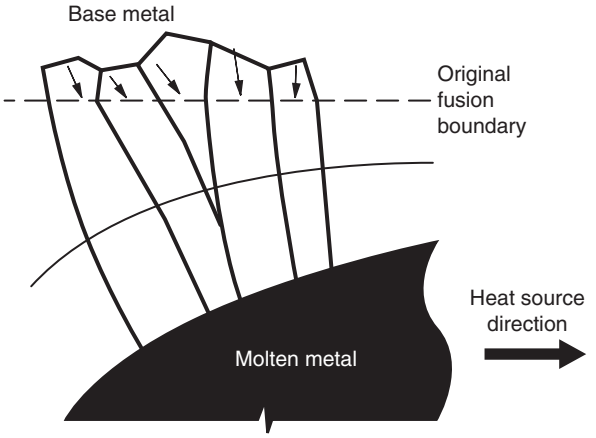


Figure 5.12 Preferential growth of crystals which have the two preferred directions of growth being coincident. Source: From Kou (1987)/with permission of John Wiley & Sons, Inc.

may be nucleated heterogeneously by fragments of dendrite tips which are broken off by convection within the molten pool, especially in the mushy zone. Both liquid and solid coexist in this region. Grains which are only loosely held together by liquid films in the mushy zone may also be swept away by convection and act as nucleation sites. Finally, new grains may be nucleated by foreign particles in the molten pool. When deliberately applied, the latter technique is called *inoculation* and is used to form a fine equiaxed grain structure. An example is the addition of titanium carbide powder as an inoculant to the molten pool during submerged arc welding of mild steel. A fine grain structure enhances the mechanical properties of the fusion zone, improving such properties as ductility and toughness. It also reduces the tendency of the fusion zone to solidification cracking (Section 5.2.2.1).

Increasing the cooling rate results in a finer solidified structure (Figure 5.7). Thus, higher traverse velocities produce finer dendritic spacings which in turn result in less segregation. They also result in a finer grain structure.

5.1.2 Zone of Partial Melting

Low-melting constituents of the base material in the region closest to the fusion boundary often melt during processing, resulting in a zone that is partially melted. This region forms a narrow band around the fusion zone (Figure 5.1). Partial melting is most pervasive along grain boundaries due to a higher concentration of impurities but may also occur within the grains. It occurs for portions of the base alloy where the temperature falls within the solidus–liquidus range or for low-melting second-phase particles or inclusions.

Under conditions of partial melting, tensile residual stresses that develop as the part begins to cool can cause separation of material, resulting in hot cracking (Section 5.2.2.1), which is normally intergranular. The impact of the partially melted region is reduced in processes where the total heat input is low, such as in laser processing. Reduction in concentration of elements that tend to increase the freezing temperature range of an alloy would also minimize damage in this zone. Finally, the low-melting constituents are normally distributed along the grain boundaries. A reduction in the average grain size, which increases the grain boundary area, would therefore reduce the concentration of the low-melting constituents, thereby minimizing their impact.

5.1.3 Heat-Affected Zone

The HAZ is the portion of the base material which does not melt, but where the temperature rise is high enough to affect the microstructure and mechanical properties of the material (Figure 5.1). The microstructure that develops in this region depends on the following:

1. The thermal cycle.
2. The type of material (its metallurgical and thermal properties).
3. The state of the material prior to processing (prior thermal and mechanical history).

For this discussion, we group the materials of interest into three main categories:

1. Pure metals.
2. Precipitation-hardening and nonferrous alloys.
3. Steels.

5.1.3.1 Pure Metals The HAZ microstructure obtained on processing a pure metal such as copper depends on whether the material is originally in the annealed state or is work-hardened. An annealed material will only experience grain growth in this area (Figure 5.13a), that is, there will be no recrystallization since rearrangement of dislocations to form new boundaries is essential for recrystallization. However, a previously work-hardened material will be recrystallized with new grains being nucleated and therefore has a fine grain structure at the outer edge of the HAZ where the temperature is just above the recrystallization temperature (Figure 5.13b). The grain size increases toward the fusion boundary, and a coarse grain structure is almost always inevitable close to the fusion boundary.

The growth kinetics of the grains is expressed as:

$$\frac{dD}{dt} = K_a \frac{1}{D} \tag{5.14}$$

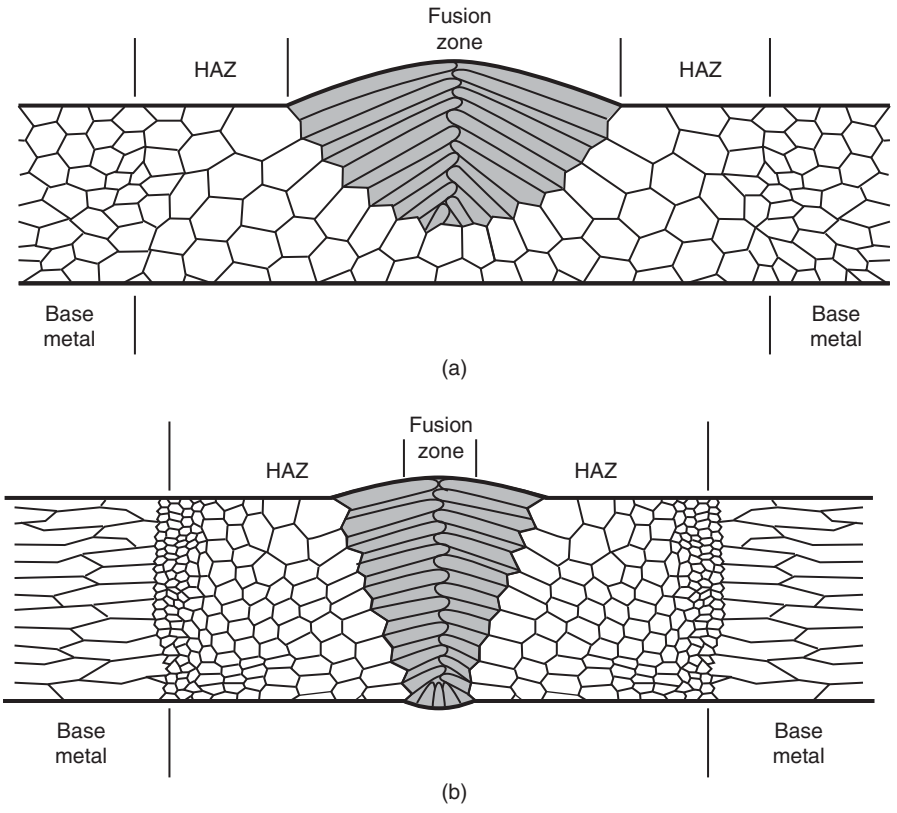


Figure 5.13 HAZ microstructure for (a) previously annealed pure metal and (b) previously work-hardened pure metal.

where D is average grain size and K_a denotes rate constant that exhibits Arrhenius temperature behavior.

Thus, Eq. (5.14) can be expressed in a more general form as:

$$\frac{dD}{dt} = K_0 e^{\left(-\frac{\Delta E_d}{k_B T}\right)} \frac{1}{D} \tag{5.15}$$

where K_0 is a diffusion constant.

For isothermal conditions, integration of Eq. (5.14) gives

$$D^2 - {D_0}^2 = K_a t \tag{5.16}$$

where D_0 is an initial average grain size.

However, since the temperature at any point in a part that is being processed varies with time, the grain size at any location at any time t is given more appropriately by:

$$D^2 = K_0 \int_0^t e^{\left(-\frac{\Delta E_d}{k_B T(t')}\right)} dt' \tag{5.17}$$

where t' is a dummy variable.

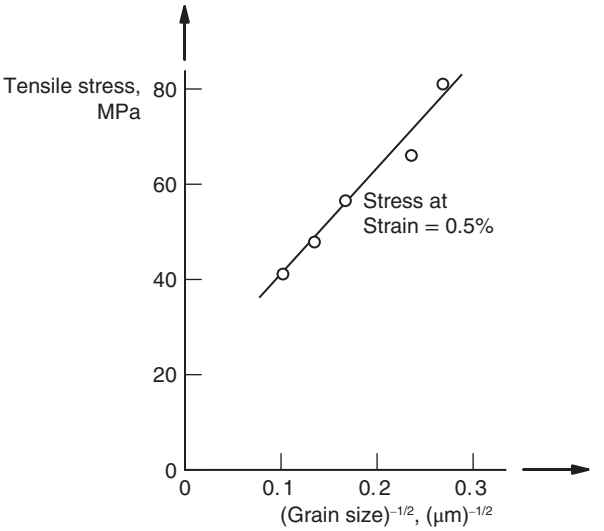


Figure 5.14 Stress variation with grain size for copper. Source: From McLean (1962). Courtesy of National Physical Laboratory, UK

The importance of grain structure stems from its influence on the mechanical properties of the component (strength and toughness). A fine grain structure results in an increase in the strength of the material. For a pure material where there are no second-phase particles, inclusions, or other microconstituents, the yield stress of the material, S_y, is related to the average grain size D by the Hall–Petch equation:

$$S_y = \sigma_i + k_y D^{-1/2} \tag{5.18}$$

where σ_i is frictional resistance of the lattice to moving dislocations and k_y is a constant.

This is illustrated in Figure 5.14 for copper. The dependence of the yield strength on average grain size is more complex for alloys, since other factors also affect the yield strength.

Example 5.3 From Figure 5.14, obtain the Hall–Petch equation for copper.

Solution:
The Hall–Petch equation is given by:

$$S_y = \sigma_i + k_y D^{-1/2}$$

where σ_i is the intercept on the S_y axis and k_y is the slope of the $S_y - D^{-1/2}$ line.

By extrapolating the line in Figure 5.14 to the S_y axis, we find the intercept to be about 20 MPa. Thus, $\sigma_i = 20$ MPa.

The slope m_g of the line is estimated from the graph to be

$$m_g = \frac{(80 - 20)\text{ MPa}}{(0.3 - 0)\ \mu\text{m}^{-1/2}} = 200\text{ MPa }\mu\text{m}^{1/2}$$

The Hall–Petch equation for this copper is thus given by:

$$S_y = 20 + 200D^{-1/2}\text{ MPa}$$

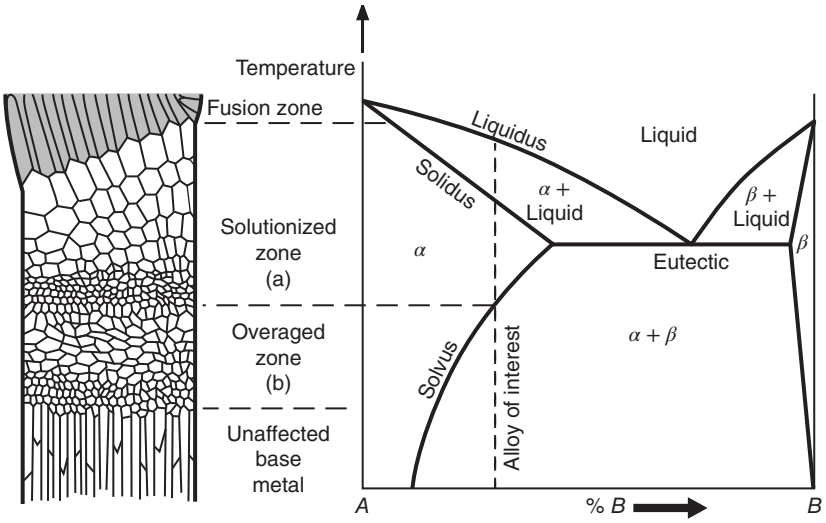


Figure 5.15 HAZ microstructure for a previously work-hardened precipitation-hardening material. Source: American Welding Society Handbook Vol. 1, 9th Edition, 2001.

5.1.3.2 Precipitation-Hardening and Nonferrous Alloys The HAZ in the case of precipitation-hardening materials, such as aluminum–copper alloys, normally consists of two regions (Figure 5.15):

a. There is the inner or solutionized zone, that is, material very close to the fusion zone, where the very high temperatures result in the material transforming into a single-phase solid solution. The new phase results in new grains being nucleated, and these grow as a result of the high temperatures. This weakens the strength of material close to the fusion line, making the strength equivalent to that of an annealed material. Irrespective of the initial condition of the material close to the fusion line, that region will be solution-treated and quenched and therefore relatively soft, of course, with the coarse grain structure persisting.
b. Slightly farther away from the fusion boundary, there is the outer or overaged zone where the material will have been heated to just a high enough temperature to result in recrystallization with an equiaxed fine grain structure if the material was previously cold-worked. Otherwise, grain growth will occur in this zone. Next to this zone, the unaffected base material will have the original structure of the material.

If the material had been previously aged, then it would become solution treated close to the fusion zone and overaged farther away. To avoid overaging, processing may be done before solution treating and aging. However, this is a difficult approach, due to the inconvenience of heat-treating large structures in furnaces and also potential distortion of the structure. Low-heat-input laser and electron beam processes reduce the overaging effect.

5.1.3.3 Steels In this section, the polymorphic property of steel is discussed in light of its effect on the HAZ grain structure. This is followed by a discussion on the HAZ microstructure of three major types of steel:

1. Low-carbon steels.
2. High-carbon and high-alloy steels.
3. Stainless steels.

Polymorphic or allotropic materials such as iron may also exhibit two distinct regions in the HAZ of a previously work-hardened material (Figure 5.16):

a. One region is due to recrystallization after transformation to the higher temperature crystal structure, for example, low austenite region. This normally occurs in the temperature range between 723 °C and 900 °C of the steel, closer to the fusion zone. In this region, the steels become either fully or partially austenitized. For the sections that are fully austenitized, new grains are nucleated, resulting in a relatively fine grain structure in the entire section. However, in the regions that are only partially austenitized, it is the newly formed austenite phase that produces a refined grain structure as new grains are nucleated, while the untransformed ferrite phase essentially maintains the original grain structure, with a possible slight growth. Thus, only partial grain refining would occur in this section. In either case, the high temperatures in the immediate vicinity of the fusion zone result in a tendency toward grain growth that may result in weakening in strength of material close to the fusion line.
b. The other region corresponds to recrystallization of the low-temperature crystal structure, that is, high ferrite region, approximately 700 °C depending on the type of steel. An equiaxed fine grain structure may be formed here. This is the outer zone next to the base material. If the steel is not work-hardened before being processed, then there will be no recrystallization, and the original grains will simply continue to grow in this region.

Low-Carbon Steels During cooling of polymorphic materials such as steels, additional phase transformation may occur in the solid state. This would often result in a new microstructure being generated that obliterates the microstructure that was generated during solidification. However, the original grain structure may exist in conjunction with new grains that are formed in the new phase.

For low-carbon, low-alloy steels, acicular ferrite may form. This has randomly oriented short ferrite needles and provides increased resistance to crack propagation, thereby increasing the fracture toughness of the steel. On the other hand, bainite reduces the fracture toughness of the material. Bainite and martensite are more likely to form with increasing carbon content and/or cooling rate.

The coarse grain structure that develops close to the fusion zone may result in a material with strength equivalent to that of an annealed material. Normally, the high cooling rates that occur close to the fusion zone, along with the coarse grain structure in that region, result in the formation of acicular ferrite along grain boundaries. Even though the low-carbon steels do not readily

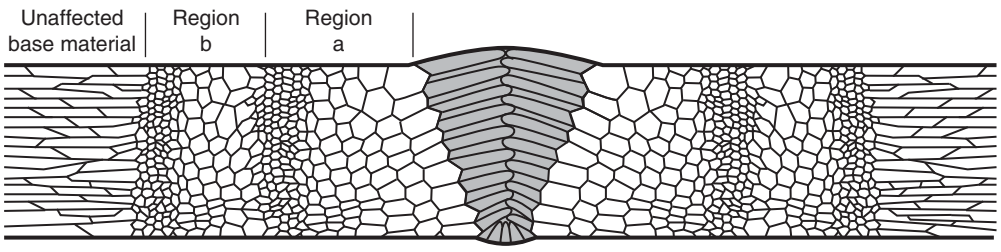


Figure 5.16 Schematic of HAZ microstructure for a previously work-hardened steel or precipitation-hardening material. Source: American Welding Society Handbook Vol. 1, 9th Edition, 2001.

form martensite, the cooling rates that accompany high-intensity (laser or electron beam) processes can produce martensite in the HAZ. The low carbon content tends to reduce the hardness of any martensite that may be formed. Furthermore, it raises the martensite start temperature such that any martensite formed is tempered during cooling (auto-tempering), thereby softening it. However, if the cooling rate is fast enough, there is insufficient time for the resulting martensite to be auto-tempered and it consequently remains as hard martensite. On the other hand, if the cooling rate is very low, the ferrite formed in the initial stages of cooling rejects carbon into the surrounding austenite which may thus subsequently transform into hard martensite since its carbon content is then high. Thus, the cooling rate for such steels have to be intermediate.

The portions of the workpiece which are heated to temperatures equivalent to that between the eutectoid temperature and the lower austenite temperature experience partial transformation, with the pearlite phase transforming to austenite which is embedded within the untransformed ferrite matrix. Subsequent cooling of the part may result in the patches of austenite transforming to hard bainite or martensite, embedded in softer ferrite matrix.

High-Carbon and High-Alloy Steels For hardenable steels such as high-carbon or high-alloy steels, the relatively high cooling rates in the region closest to the fusion line usually transform material in this region to martensite.

Hardenability refers to the ease with which martensite forms. It does not refer to the hardness or strength of the steel. Plain carbon steels, for example, have low hardenability and as such, only very high cooling rates can produce martensite in them. On the other hand, even air cooling may result in the formation of martensite in some high-carbon or high-alloy steels.

Since different alloying elements affect the hardenability of a steel differently, a simple way of estimating the hardenability or susceptibility of a given steel to cracking is to determine its carbon equivalent (CE). This combines the impact of all the alloying elements into a single carbon fraction that provides, in a nutshell, an estimate of the material's hardenability. The carbon equivalent relationship depends on the alloying elements, and there are various forms of it. A commonly used expression is

$$\mathrm{CE} = C + \frac{\mathrm{Mn}}{6} + \frac{\mathrm{Cr} + \mathrm{Mo} + V}{5} + \frac{\mathrm{Ni} + \mathrm{Cu}}{15} \tag{5.19}$$

The variables are expressed in percentage by weight. Preheating or postweld heat treatment is often not necessary for materials with a carbon equivalent less than 0.35%. Preheating is normally required when the carbon equivalent is between 0.35 and 0.55%, while for a CE greater than 0.55%, both pre- and postheat treatments are necessary.

Stainless Steels The corrosion resistance of iron increases as its chromium content increases, since the chromium enables a thin, protective surface layer of chromium oxide to be formed on the surface when the material is exposed to oxygen. Stainless steels are essentially iron–chromium alloys which are corrosion-resistant to an extent that depends on the alloying elements of the steel. Common alloying elements include nickel, carbon, and molybdenum, and the types and amounts of theses alloying elements determine the type of the stainless steel. There are three principal types:

(1) Ferritic stainless steels.
(2) Austenitic stainless steels.
(3) Martensitic stainless steels.

Each of these behaves differently in the HAZ, and they are briefly described in the following paragraphs.

Ferritic Stainless Steels This type of stainless steel (Fe-Cr) has a relatively low carbon content (less than 0.12% C) and a high chromium content (12–30% Cr). With this alloy content, the material remains partially or completely ferritic at high temperatures. An example is Type 430. The high chromium content gives it high corrosion resistance, even though it also makes it relatively more expensive.

Ferritic stainless steels experience severe embrittlement when held in the temperature range of 450–550 °C for some time, resulting in a loss in toughness and increase in hardness. The only way to remove the embrittlement involves reheating the material to a higher temperature and quenching through this temperature range. As a result, this type of stainless steel is not appropriate for parts that need to be welded, since the portion of the part that is heated to this temperature range will have poor properties.

Another problem encountered when ferritic stainless steels are processed with a heat source is the extensive grain growth that occurs, since there is no phase change to produce a refined grain structure, as the material is heated in the solid state. Thus, in this material, grain refinement can only be induced by cold working followed by recrystallization, which is not convenient for parts that are already processed. Finally, ferritic stainless steels that contain more than 16% Cr also encounter sensitization (see the next subsection on *Austenitic Stainless Steels*) when heated above 925 °C. This type of sensitization is not effectively subdued by lowering the carbon content or rapid quenching from above the sensitization temperature, as is commonly done for austenitic stainless steels. However, additions of niobium and/or titanium can suppress the sensitization problem.

Austenitic Stainless Steels Stainless steels with a relatively high chromium content and significant amounts of nickel (Fe-Cr-Ni), typically 18% Cr and 8% Ni stabilize the austenite phase to the point that it still exists even at room temperature. Thus, such alloys cannot be hardened by heat treatment and are only hardened by cold working. An example is Type 301. When this type of stainless steel is cold-worked, the unstable austenite is transformed into martensite, significantly increasing its strength.

When austenitic stainless steels are heated between 425 and 800 °C, carbon tends to react with chromium to form chromium carbide, $Cr_{23}C_6$, and the reaction is most rapid around 650 °C. Thus, in the regions of the HAZ which are heated to this temperature range, chromium carbide is precipitated, usually at the grain boundaries. This phenomenon is called *sensitization* and induces corrosion (Figure 5.17), for two reasons:

1. An electrolytic cell is formed at the grain boundaries.
2. The material loses corrosion resistance along grain boundaries due to a significant reduction in the amount of free chromium.

Sensitization can be prevented by:

1. Reheating the part that has been processed to about 1100 °C to redissolve the chromium carbide and cooling fast to prevent the carbide from precipitating again. However, distortion

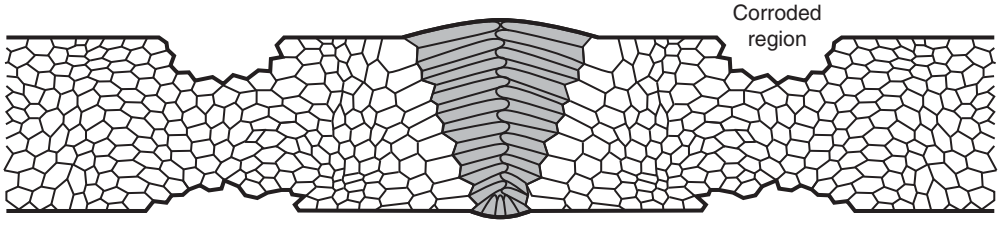


Figure 5.17 Corrosion in an austenitic stainless steel after welding, as a result of sensitization.

of the part that sometimes results from quenching often makes this treatment undesirable. Furthermore, if the structure being processed is cumbersome, then the heat treatment process becomes unfeasible.

2. Reducing the carbon content of the alloy to say 0.03% C so that little chromium is tied up as chromium carbide. An example is Type 304, which is easy to weld.
3. Adding niobium or titanium that has a greater affinity for carbon than does chromium, thus having all the available carbon tied up leaving the chromium free in solution. Examples are Type 347 which contains niobium, and Type 321 which contains titanium. However, the HAZ of Types 321 and 347 can still suffer attack if subjected to severe corrosive conditions after heating above 1100 °C and 1300 °C, respectively. This is because the Ti and Nb carbides go into solution when held at those temperatures, resulting in subsequent chromium carbide precipitation along grain boundaries in a narrow zone close to the fusion boundary. The resulting corrosion is referred to as knife-line attack. This type of problem can be eliminated by limiting the carbon content to less than 0.03% C.

Due to the relatively low solubility of sulfur in austenite, austenitic stainless steels tend to have a high susceptibility to hot cracking (Section 5.2.2.1). This crack sensitivity is minimized by introducing small amounts of δ-ferrite, say 5–10%, to form a duplex structure containing a mixture of ferrite and austenite, since ferrite has a greater solubility for sulfur.

Elements which tend to stabilize the austenite phase include carbon, nickel, manganese, and nitrogen, while those which promote the formation of ferrite are chromium, niobium, aluminum, silicon, and molybdenum.

Martensitic Stainless Steels Martensitic stainless steels normally have chromium in the range 12 to 18%, and carbon in the range 0.15 to 1.20%. An example is Type 410, and most of them form martensite on air-cooling from the austenite. They usually encounter severe temper embrittlement when heated in the temperature range 370–600 °C.

5.2 DISCONTINUITIES

Discontinuities are normally introduced into a part during processing. As defects (i.e. result in rejectable condition), they are undesirable as they reduce the mechanical properties of the material. They thus influence the safety and performance of the part. There are various types of discontinuities, and the principal ones include porosity and cracking, which are metallurgical; incomplete penetration, lack of fusion, and undercutting, which are process- and technique-related (Figure 5.18). It is sometimes also useful to identify discontinuities by geometry, that is, whether they are planar (two-dimensional) or volumetric (three-dimensional). The planar types include cracks, lack of fusion, and incomplete penetration, and these tend to have a much greater damaging effect than do the volumetric types. The volumetric types include porosity. The closer a given discontinuity is to the surface, the more detrimental it can be to the integrity of the part. Furthermore, if the discontinuity is oriented such that the applied stress tends to open it up, then its damaging effect is greater than one that is oriented otherwise. Temperature, stress state, time, and service environment do have significant influence on defects. Modes of failure include plastic deformation, fatigue, fracture, corrosion, and creep.

In the following sections, the principal discontinuities are discussed.

5.2.1 Porosity

Porosity is the existence of gas pores or pockets within the fusion zone (Figure 5.19) and may be defined as cavity-type discontinuities (voids) formed by gas entrapment during solidification.

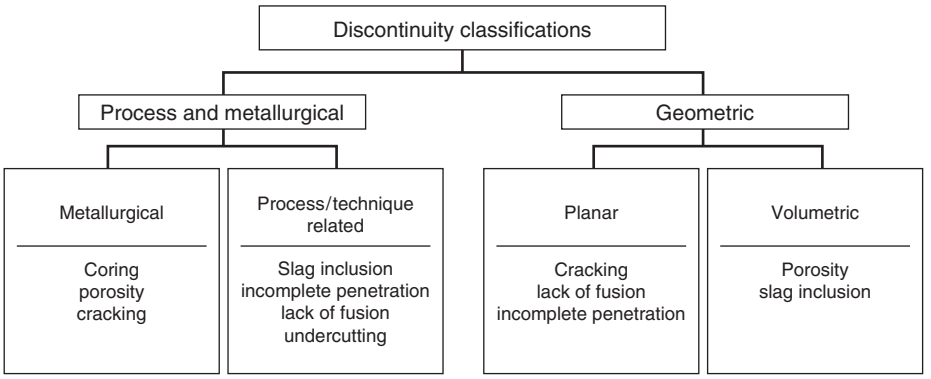


Figure 5.18 Classification of common discontinuity types.

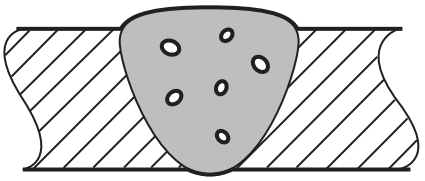


Figure 5.19 Porosity in the fusion zone.

It basically arises from the absorption of gases into the melt (droplet or pool) and subsequent release of part of the gases by the molten metal during cooling due to reduced solubility in the solid. Surface contamination (e.g. oils, moisture, and paint) is a major source of the gases that cause porosity. Gases may also be released by chemical reactions in the molten metal. Oxygen, hydrogen, and nitrogen are the principal sources of porosity and are absorbed atomically.

The rate of solidification of the molten pool has a pronounced effect on porosity formation. Porosity formation proceeds by nucleation and growth just as occurs during solidification. The bubbles that escape to the surface if the growth rate exceeds the solidification rate, and they are trapped as pores if the reverse is true. Porosity is usually most severe in the initial stages of a process, since the conditions are not steady during that period.

Oxygen may be introduced into the molten metal by oxides on filler material or on the base material. It may also be introduced from the atmosphere. The presence of oxygen in the molten pool has two effects. One involves oxidation of the fused metal to produce material with poor properties. The other involves a reaction with carbon in the case of steel to form carbon monoxide which is trapped in the fusion zone. The influence of oxygen may be minimized by adding deoxidants, for example, manganese, silicon, or aluminum to any filler metal used. The oxides so produced are much less damaging than the porous structure that would otherwise result. However, excess deoxidants may reduce the mechanical properties of the fusion zone. The oxides formed when aluminum is used as a deoxidant tend to stay in the fusion zone and give it poor mechanical properties. The addition of silicon and manganese along with aluminum tends to produce oxides which float and are therefore removed with slag.

Hydrogen may, in addition to the porosity, also makes the fusion zone and HAZ brittle. It is considered as the principal cause of porosity when welding steels and aluminum. This is because there is a significant difference in the solubility of hydrogen in the molten metal at high temperatures and its solubility at the melting temperature. The excess hydrogen may be released in the molten pool

as it cools. Hydrogen may enter the molten pool from the atmosphere. Moisture absorbed by oxides on the surfaces of filler rod and workpiece and humidity in the air are often the main sources of hydrogen and oxygen. The water vapor dissociates to form hydrogen and oxygen. Thus for highly reactive materials like aluminum and magnesium, any filler materials to be used need to be protected from the atmosphere until used or cleaned before being used.

Brasses are difficult to weld due to volatilization of the zinc alloying element, which may result in porosity and also affect visibility. Copper oxide reacts with hydrogen to release steam, which may induce porosity in the fusion zone unless the copper is deoxidized by adding phosphorus and/or silicon, manganese, titanium, and aluminum.

The most suitable traditional method for detecting porosity after processing is the radiographic technique, followed by the ultrasonic inspection method. Liquid penetrant, magnetic particle, and visual inspection are useful only when the pores appear on the surface. Porosity levels up to about 7% by volume generally do not have any significant influence on the tensile properties of the part. However, the ductility tends to reduce slightly with increasing porosity, the effect being more pronounced as the strength of the material increases. Failure in fatigue is influenced by porosity only when any reinforcement that is present is removed, and surface porosity tends to be more damaging than buried ones. A reinforcement is the portion of the fusion zone that projects above the workpiece surface. More often than not, it results from filler material that is added to the molten pool.

5.2.2 Cracking

In this section, we discuss the basic concepts of cracking that occurs during and after processing. Cracking is the most damaging of all the defects. The principal forms of cracking are as follows:

1. Hot cracking.
2. Liquation cracking.
3. Cold cracking.

Cracking may occur in either the fusion zone, the HAZ, or both (Figures 5.20 and 5.21). The crack location, in the case of carbon steels, is influenced by the hardenability of both the fusion zone and base metal. The cracks may be in the form of centerline cracks if the fusion zone and base metal compositions are about the same and the workpiece is not severely restrained. This is because the highest cooling rate occurs along the centerline. On the other hand, in situations where a filler metal is used and its carbon content is lower than that of the base metal, the overall carbon content and thus hardenability of the resulting fusion zone are lower than that of the base metal. The coarse-grained region of the HAZ then tends to be more susceptible to crack formation, resulting in cracks appearing in the HAZ instead.

Cracks could be formed either during the process, that is, soon after solidification (hot cracking) or after the bead has cooled to room temperature (cold cracking), and may continue for several hours, days, or even longer. Hot cracking is more prevalent in the fusion zone, whereas cold cracking occurs more commonly in the HAZ. The principal factor inducing cracking is the residual stresses that are set up as the fusion zone contracts during cooling. The stresses are tensile in the fusion zone and gradually change to balancing compressive stresses in the surrounding material.

Cracks are unfortunately difficult to detect using conventional nondestructive evaluation techniques due to its planar form. Ultrasonic testing is the best method for off-line detection of embedded cracks, but when other forms of discontinuity exist, it becomes difficult distinguishing between them. For real-time or online detection, acoustic emission, audible sound, and infrared/ultraviolet detection have been found to be effective.

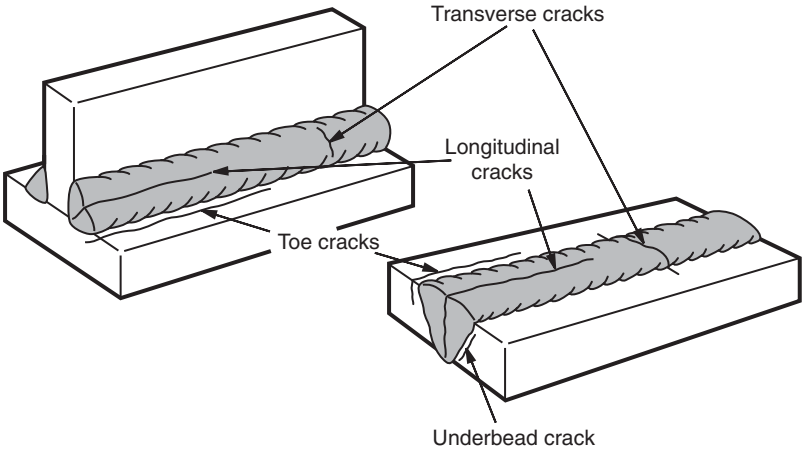


Figure 5.20 Schematic illustration of weld cracks. Source: American Welding Society Handbook Vol. 1, 9th Edition, 2001.

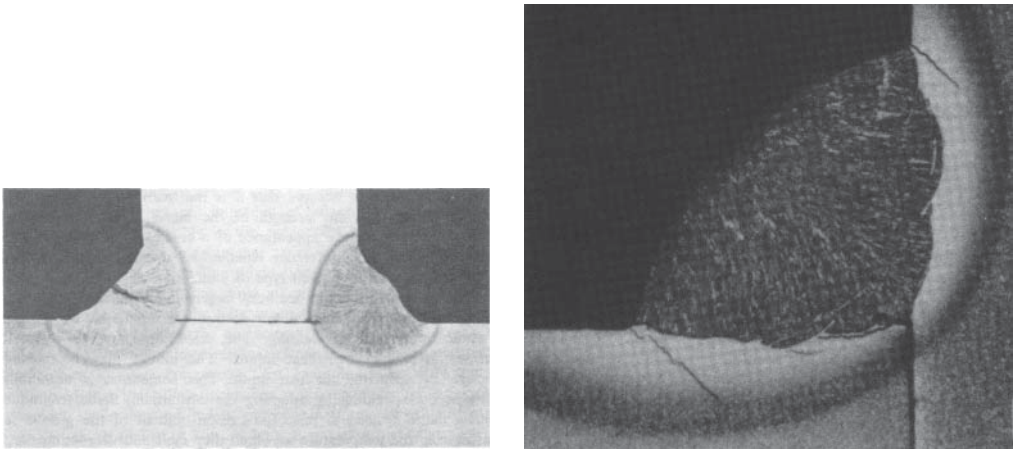


Figure 5.21 Pictorial illustration of weld cracks. (a) Hot crack in a fillet weld. (b) Cold crack in the heat-affected zone of a fillet weld. Source: Kou (2003)/with permission of John Wiley & Sons, Inc.

5.2.2.1 Hot Cracking Hot cracking usually occurs in the fusion zone either during or immediately after solidification, in the temperature range of about 200–300 °C below the melting temperature. It results primarily from low-melting constituents that are capable of wetting the grain boundaries and which form either interdendritic or intergranular liquid films. The segregation of these films results in localized regions of weakness along the grain boundaries and between dendrites which rupture under the tensile residual stresses that develop as the part cools down. This produces interdendritic or intergranular cracks, that is, the cracks occur in between grains or dendrites, or along their boundaries, instead of occurring within the grains or dendrites themselves. They often develop along the centerline also. The incipient cracks may heal if there is adequate molten metal near the cracks to backfill them. Due to their high-temperature occurrence, hot cracks exposed to the surface can often be identified by an oxide coating that discolorates the crack surfaces. Whether a crack is

formed or not depends on the type of intergranular films formed, the microstructure (that is, whether it is coarse or fine), and the residual stresses developed in those regions. Fine equiaxed grains tend to have greater resistance to hot cracking than do coarse columnar grains, since the low-melting constituents are then distributed over a wider surface area, and thus are in lower concentrations. Contaminants such as oils that are left on the base metal may also contribute to hot cracking.

In steels, non-metallic alloying elements such as sulfur, phosphorus, and boron that tend to promote the formation of low-melting integranular films (which are capable of wetting the grain boundaries) make the material more susceptible to hot cracking. Sulfur forms iron sulfide, FeS, which dissolves in the molten steel and increases the freezing range between the solidus and liquidus considerably. Furthermore, the surface tension between the molten iron sulfide and the solid grains is low, resulting in wetting of the grain boundaries, that is, the formation of a liquid film. With traces of low-melting phases rich in sulfur still existing in liquid form at relatively low temperatures when significant residual stresses are set up, hot cracking can easily occur. Hot cracking will generally not occur if the sulfur and phosphorus content is less than 0.06%. An increase in the carbon content also increases the cracking tendency, as it increases the freezing range. Nickel has a similar effect, especially in concentrations above 3%. Thus, fusion welds are difficult when the nickel content exceeds 4%.

The solubility of sulfur in austenite is less than in ferrite. There is therefore a greater tendency for the liquid grain boundary film to form and cause cracking in austenitic steels than in ferritic steels, since the sulfur is more likely to segregate and form weak regions in the former. Addition of austenite stabilizers such as carbon, nickel, and copper to steels thus makes them more crack-susceptible.

Manganese tends to make the sulfides globular and thereby prevents cracking when used in the right amounts. This is because it does not dissolve in molten steel or increase the freezing range and is therefore not dispersed as an intergranular film. As the carbon content of the steel increases, so must the manganese-to-sulfur ratio to prevent hot cracking. However, at high carbon contents, say greater than 0.3% C, higher ratios of manganese to sulfur are no longer effective.

Hot cracking in nonferrous alloys may occur for compositions that fall in regions of the phase diagram with a high temperature range between the liquidus and solidus lines, that is, a long freezing range. Under such circumstances, there is a greater tendency for the formation of low-melting phases.

The sensitivity of the aluminum–silicon alloy (4000 series), for example, to solidification cracking increases with increasing silicon composition, peaking at about 0.5% silicon and then decreases with further increases in the silicon content. Aluminum–magnesium alloys (5000 series) are also sensitive to hot cracking. In the case of magnesium, zinc (4–6%) and calcium are the alloying elements that result in the highest susceptibility to hot cracking, while manganese, aluminum, and zirconium alloying elements do not present significant hot cracking problems. The presence of contaminants such as oil and paint on the surface of nickel workpieces are often the source of low-melting constituents which may result in hot cracking or liquation cracking in welding nickel and its alloys. Adequate cleaning of such materials is thus essential for defect-free welding. The high affinity of the reactive metals such as titanium and zirconium to oxygen and nitrogen makes them difficult to weld, since they result in embrittlement of the joint and may also induce porosity. Thus, with such materials, it is often necessary to introduce shielding gas on both sides of the joint.

The tendency for hot cracking may be minimized by:

1. Adequate cleaning of the workpiece to eliminate surface contaminants such as oil and paint.
2. Improved design to minimize constraint effect and thus residual stresses.
3. Using material with low cracking sensitivity.

5.2.2.2 Liquation Cracking Sometimes, complex carbides or low-melting intermetallic compounds that result from segregation may melt to produce local grain boundary films or regions of

localized brittleness in the zone of partial melting during processing. This is especially likely in the temperature range between the solidus and liquidus curves, even though the bulk of material in that region may still be solid. Under sufficiently high residual stresses, cracks may open up in these regions, causing liquation cracking. Liquation cracking may result from sulfur and phosphorus-forming grain boundary liquation in steels, or niobium-forming eutectic patches. In the former case, the addition of manganese does not alleviate the problem. The only solution is to limit the amounts of sulfur and phosphorus in the steel to very low amounts. For aluminum, the Al-Mg-Zn alloys or duralumin are most susceptible to this form of cracking.

Hot cracking involves material that is initially completely molten and is thus common in the fusion zone, whereas liquation cracking involves partial melting, and is most common in the HAZ or zone of partial melting.

5.2.2.3 Cold Cracking The high cooling rates that occur during processing may result in brittleness in both the fusion zone and portions of the HAZ. The brittleness is even more likely in steels when martensite is formed. These factors, along with the presence of hydrogen, promote crack formation in steels, which may start soon after solidification and continue for some days afterward. This type of crack normally occurs after the material has cooled to relatively low temperatures and is thus referred to as cold cracking.

Cold cracking is usually transgranular in nature but may sometimes also be intergranular. It may appear at the surface of the part, for example, toe cracks, or remain hidden beneath the fusion zone as underbead cracks (Figures 5.20 and 5.21). It is generally induced by the following:

1. Hardening due to rapid cooling (martensite formation).
2. Residual stresses that are set up in the part.
3. Embrittlement by hydrogen (is enhanced by a moist atmosphere).
4. Existence of hot cracks.

Residual stresses are often introduced in processes that involve a heat source. The formation of cold cracks thus depends primarily on the cooling rate (formation of martensite) and the presence of hydrogen, for steels. Without hydrogen, cold cracking is less likely to occur in steels. Materials with high martensite start and martensite finish temperatures have a lower susceptibility to cold cracking, since there is a drastic reduction in the solubility of hydrogen when transformation occurs from austenite to martensite. Also, at the higher temperatures, there is a greater tendency for the supersaturated hydrogen to diffuse out. Furthermore, the higher transformation temperature enables the marternsite formed to be tempered to some extent, thereby reducing the associated hardness and residual stresses.

Increasing the hydrogen content of the shielding gas has been shown to directly increase the extent of cold cracking. Hydrogen may be introduced through moisture on electrodes and workpiece or from the atmosphere. Its influence is twofold:

1. Causing porosity.
2. Embrittlement.

Even though the presence of hydrogen is almost always essential for cold cracking to occur, a high-carbon high alloy steel that is highly restrained can result in cracking in a virtually hydrogen-free atmosphere.

Cold cracking may be prevented by:

1. Preheating to reduce the cooling rate (Chapter 4) and thus minimize the formation of martensite in steels (preheating also helps to remove moisture).

2. Postweld heat treatment (about 550–650 °C for steels) to soften any martensite formed and help hydrogen diffuse out (also helps remove residual stresses).
3. Improved design to minimize constraint effect and thus residual stresses.
4. Using material with low cracking sensitivity, for example, low-carbon steels.

Preheating is about the most commonly used method for preventing cracking when processing high hardenability materials. Conventional preheating and postweld heat treatment involve either heating the workpiece in a furnace before processing or using an induction or flame heater. For the latter two methods, the heating is often applied locally. In the case of furnace heating, the entire workpiece is heated to the preheat temperature, which is about 250 °C for a number of steels. These methods are often either too expensive or impractical to use. For laser welding, a more convenient approach may involve using a second laser beam to either preheat the joint or postweld heat treat the fusion zone during the process (Section 10.7.1.1).

5.2.3 Lack of Fusion

When the heat supply is not adequate, it may result in lack of fusion of the molten region (Figure 5.22), where there is no fusion between the fusion zone and base material. Lack of fusion may also be caused by flooding of molten metal in front of the heat source. It is more commonly observed in processing pure copper because of the high thermal conductivity which results in most of the heat supplied being lost to the surrounding material, but can be minimized or eliminated by preheating the workpiece before processing.

Ultrasonic inspection is most suited for identifying lack of fusion, since this type of discontinuity can be either planar or volumetric. Radiographic inspection will detect the volumetric ones but not the more planar type. It is generally considered that up to 10% lack of fusion will not have a significant effect on the tensile properties. However, for a part loaded in fatigue and which has the reinforcement removed and no postweld heat treatment, a 10% reduction in area due to lack of fusion is known to reduce the fatigue strength by about 50%. Lack of fusion is most damaging at a weld face or root.

5.2.4 Incomplete Penetration

Incomplete penetration occurs when the filler and base metals do not fuse at the weld root (Figure 5.23). The discontinuity is referred to as such only when the penetration is less than the

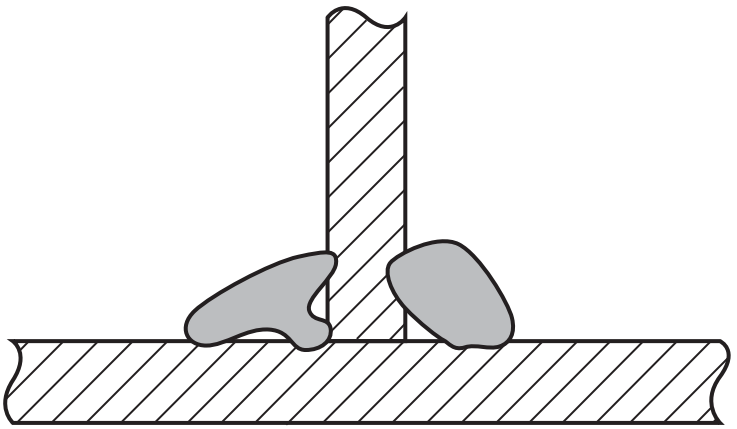


Figure 5.22 Lack of fusion.

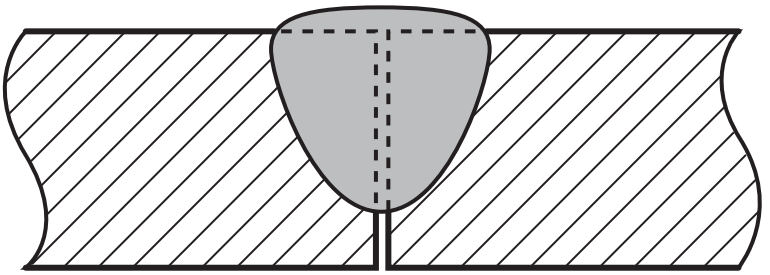


Figure 5.23 Incomplete penetration.

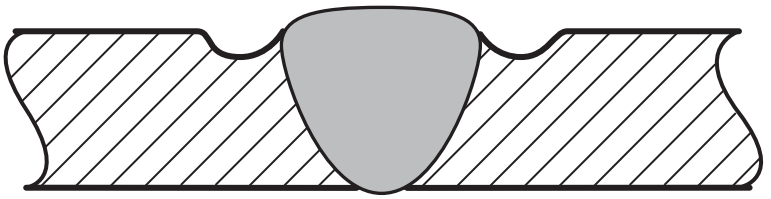


Figure 5.24 Undercut.

design specification. It is usually caused by low heat input. The inspection techniques and effect on properties are essentially the same as for lack of fusion. Incomplete penetration reduces the load-carrying capacity of the part.

5.2.5 Undercut

This is the portion of the base metal that is melted but not filled (Figure 5.24) and forms a groove. It is usually a region of high stress concentration at the weld toe. High processing speed is a common cause of undercutting, since under these conditions, the filler metal being deposited does not have sufficient time to flow to the toe. Conditions that induce a marked convex contour of the fusion zone, such as material with high surface tension, also enhance undercutting. These tend to draw the molten metal in the molten pool upward and to the rear, away from the sides of the bead, with the result that a positive meniscus develops and the molten metal fails to wet the sides of the joint. It is generally not harmful unless the part is subjected to severe fatigue loading conditions.

5.3 SUMMARY

In this chapter, we discussed the basic microstructural changes that occur in a material as a result of laser processing and were broken down into two main sections:

- Process microstructure.
- Discontinuity formation.

Under process microstructure, there are four major zones that result after processing: the fusion zone, zone of partial melting, HAZ, and base metal. Solidification in the fusion zone involves nucleation and growth. The concept of epitaxial solidification, which is common with laser processes that

result in melting without the addition of filler material, involves solidification without nucleation. In conventional solidification, a critical radius and a critical free energy change are necessary to form a stable nucleus. The two common modes of nucleation are homogeneous and heterogeneous nucleation. After a stable nucleus is formed and begins to grow, the structure that results in the fusion zone may be one of four types: planar, cellular, dendritic columnar, or dendritic equiaxed. The microstructure varies from a planar to a cellular or dentritic structure as one moves from the fusion boundary toward the centerline due to changes in solidification rate and temperature gradient. In addition to the general structure, the grain orientation also depends on the direction of maximum temperature gradient and the preferred crystallographic growth direction. A higher cooling rate produces a finer grain structure and also a smaller spacing between the dendrite arms, which in turn results in greater strength, better ductility, and better toughness. A finer grain structure can also be produced by inoculation, a process that involves the introduction of fine particles into the molten pool.

The zone of partial melting, which is next to the fusion zone, results primarily from low-melting impurities. Next to it is the HAZ, where no melting occurs, but the microstructure is changed by the heat flow. If the material is previously cold-worked, recrystallization occurs in this zone, resulting in a fine grain structure that increases in size toward the fusion boundary. Precipitation-hardening alloys and allotropic materials such as steel may exhibit two recrystallized regions in the HAZ. If the material is not cold-worked, then the original grains grow in size toward the fusion boundary. Ferritic and martensitic stainless steels may encounter severe embrittlement during processing, while austenitic steels may be subjected to sensitization.

Finally, the principal types of discontinuities that can occur during laser processing were discussed, specifically porosity, cracking (hot cracking, liquation cracking, and cold cracking), lack of fusion, incomplete penetration, and undercutting. Cracking is quite likely in high-carbon and high-alloy steels unless they are preheated before being processed.

APPENDIX 5.A

List of symbols used in the chapter

Symbol	Parameter	Units
ΔE_d	Activation energy for diffusion	J
ΔG_c	Critical free energy change necessary to form stable nuclei	J
ΔG_v	Bulk free energy change per unit volume	J/m^3
k_y	Hall–Petch equation constant	$N/m^{3/2}$
K_a	Rate constant that exhibits Arrhenius temperature behavior	m^2/s
K_0	Rate constant for diffusion	m^2/s
$\dot{n}$	Rate of nucleation	$/m^3$ s
n_v	Number of nuclei per unit volume	$/m^3$
r	Radius of a nucleus	m
r_c	Critical nucleus radius	m
s	Number of sites available for nuclei formation	—
t'	Dummy variable of integration	s
V_m	Volume of a molecule	m^3
γ_{nl}	Surface energy per unit area between the nucleus and liquid	J/m^2

Symbol	Parameter	Units
γ_{ns}	Surface energy per unit area between the nucleus and solid base metal	J/m^2
γ_{sl}	Surface energy per unit area between the solid base metal and liquid	J/m^2
ϕ	Wetting angle	°
ψ	Angle between traverse velocity and direction of solidification	°
σ_I	Frictional resistance of the lattice to moving dislocations	N/m^2
σ_y	Yield strength of a material	N/m^2

PROBLEMS

5.1. Most steels derive their strength from quenching, but other materials (including most aluminum alloys) are strengthened by precipitation-hardening. During the aging process, the strength of the part varies with time as shown in the Figure P5.1. If a precipitation-hardening material is welded, show the relative strengths in the three regions of the weldment: A (weld bead), B (heat-affected zone), and C (base material) in the figure if the material was ideally aged before welding.

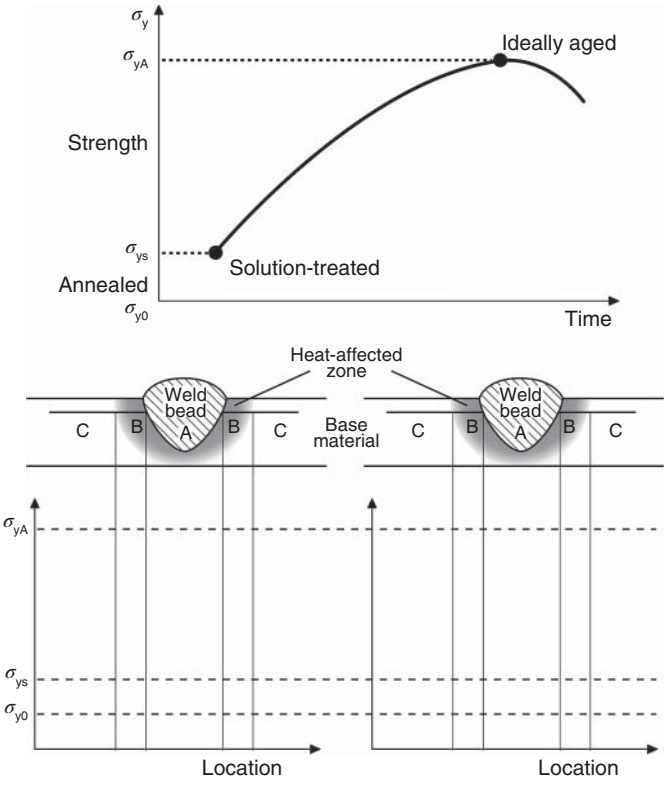


Figure P5.1 Aging curve and strength variation in a processed zone.

(a) Immediately after processing
(b) After subsequent artificial aging.

5.2. Given that the total free energy change associated with the formation of a spherical cap solid nucleus is given by:

$$\Delta G = \left[\frac{4}{3}\pi r^3 \Delta G_v + 4\pi r^2 \gamma_{nl}\right] f(\varphi)$$

show that the critical free energy change necessary for the formation of a stable nucleus is

$$\Delta G_c = \frac{16\pi {\gamma_{nl}}^3}{3(\Delta G_v)^2} f(\varphi)$$

5.3. When a material that transforms from one phase to another at an equilibrium temperature is cooled to that temperature, transformation will begin. At that temperature, the change in free energy is zero. If it is cooled below the equilibrium temperature, the change in free energy is negative and there is a greater driving force for the transformation to proceed. The lower the temperature, the greater will be the driving force. Thus, one would expect the shape of the top part of a T–T–T diagram (Figure P5.3) (above the nose) to be logical, since the lower the temperature, the sooner will the transformation begin.

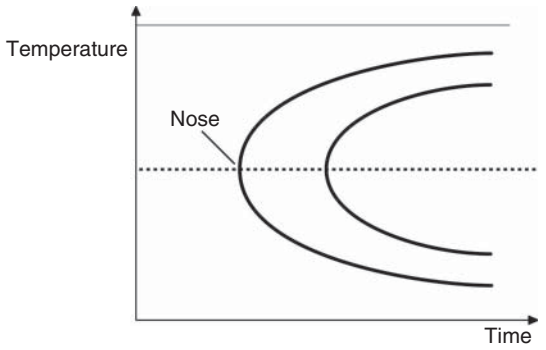


Figure P5.3 Typical time temperature transformation curve.

Why then does the transformation take longer at lower temperatures below the nose of the T–T–T diagram?

5.4. The variation of strength with time during aging at 200 °C of an aluminum alloy which was solution treated at 450 °C and then quenched to room temperature is shown in Figure P5.4. Sketch the curve for aging at 250 °C.

(a) Name two main differences between the two curves.
(b) Why are the two curves different?
(c) Sketch the aging curve for a pure aluminum piece if it were subjected to the same treatments.
(d) Is it different? Explain your solution.

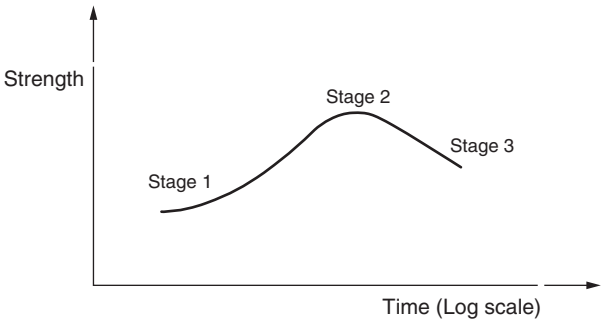


Figure P5.4 Variation of strength with time during aging.

5.5. (a) Determine the heat of fusion per atom for copper if its atomic weight is 0.0635 kg/mol.
(b) Determine the free energy change necessary to form a critical nucleus at a temperature that corresponds to 100 °C of supercooling.
(c) Determine the free energy change necessary if the supercooling is 250 °C.
(d) Comment on the probability of observing homogeneous nucleation in both (b) and (c).

5.6. The variation of yield stress with grain diameter for different alloys is shown in Figure P5.6. Obtain the Hall–Petch equation for Fe-N from information given in the graph.

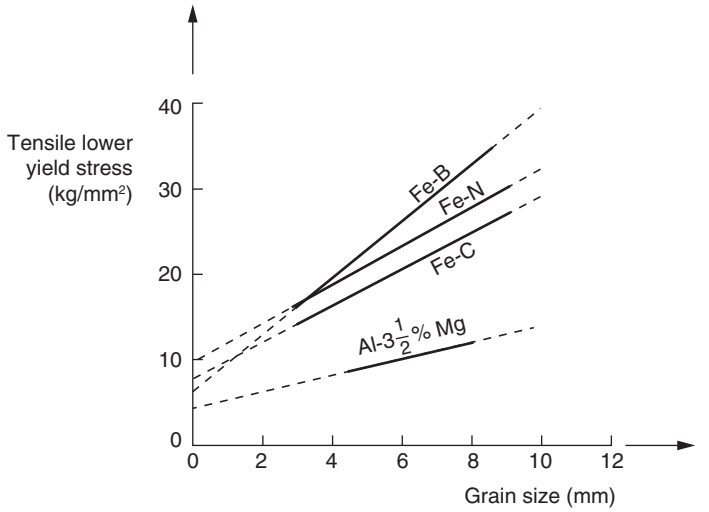


Figure P5.6 Stress variation with grain size for different alloys.

5.7. Two plates of silver are welded together using a laser beam. If the welding conditions are such that solidification of the molten pool starts with homogeneous nucleation, determine the critical radius that will result in the formation of a stable nucleus at a temperature that is 250 °C below the melting temperature.

5.8. Determine the critical free energy change necessary for forming the stable nucleus in Problem 5.7.

5.9. Determine the number of stable nuclei per unit volume for the conditions in Problem 5.7 if the density of nucleation sites is 6×10^{48} per m^3.

5.10. Calculate the nucleation rate for the conditions in Problem 5.7. Assume the radius of a silver atom is 0.144 nm (van Vlack, 1975, p. 95) and the activation energy for self-diffusion in silver is 44.1 kcal/mol (Shewmon, 1983, Table 2–2, p. 66).

5.11. (a) A part made of nickel has its surface modified by surface melting (Chapter 11) using a laser. Assuming solidification of the molten material occurs primarily by homogeneous nucleation, plot the variation of the critical nucleus size and critical free energy necessary for nucleation with the degree of supercooling.

(b) Now, if instead of just surface melting, the surface is modified by cladding, and the filler metal used results in a wetting angle of 30°, again plot the variation of the critical nucleus size and critical free energy necessary for nucleation with the degree of supercooling.

(c) How does supercooling affect the nucleation process? In what way does the use of a filler material affect the process?

5.12. (a) A tin plate is surface-melted using a laser beam. Determine the temperature to which the molten pool was supercooled if the critical radius necessary for the formation of a stable nucleus is found to be 0.9 nm.

(b) If, instead of surface melting, cladding is performed, resulting in a wetting angle of 30°, calculate the corresponding change in critical free energy.

Assume the following properties of tin:

Average density = 7168 kg/m^3
Heat of fusion = 59.5 kJ/kg
Melting temperature = 231.9 °C = 504.9 K

Surface energy per unit area for tin = 0.0545 J/m^2.

BIBLIOGRAPHY

Barrett, C.R., Nix, W.D., and Tetelman, A.S. (1973). *The Principles of Engineering Materials*. Englewood Cliffs, New Jersey: Prentice Hall.

Easterling, K. (1983). *Introduction to the Physical Metallurgy of Welding*. London: Butterworths.

Felbeck, D.K. and Atkins, A.G. (1996). *Strength and Fracture of Engineering Solids*, 2e. Englewood Cliffs: Prentice Hall.

Friedman, P.A. and Kridli, G.T. (2000). Microstructural and mechanical investigation of aluminum tailored-welded blanks. *Journal of Materials Engineering and Performance* 9: 541–551.

Granjon, H. (1972). *Cracking and Fracture in Welds*, Japan Welding Society, IB1.1.

Guy, A.G. (1971). *Introduction to Materials Science*. New York: McGraw-Hill.

Hughel, T.J. and Bolling, G.F. (ed.) (1971). *Solidification*. Metals Park, OH: American Society for Metals.

Hosford, W. (2005). *Physical Metallurgy*. Boca Raton, FL: Taylor & Francis.

Jenney, C.L. and O'Brien, A. (ed.) (2001). *Welding Handbook*, 9e. Miami, FL: American Welding Society.

Kou, S. (2003). *Welding Metallurgy*, 2e. New York: Wiley.

Lancaster, J.F. (1980). *Metallurgy of Welding*, 3e. London: George Allen & Unwin.

Linnert, G.E. (1968). *Welding Metallurgy*. Miami, FL: American Welding Society.

Lundin, C.D. (1984). Fundamentals of weld discontinuities and their significance. *WRC Bulletin* 295: 1–33.

Mallett, M.W. and Rieppel, P.J. (1946). Arc atmospheres and underbead cracking. *Welding Journal* 25: 748s–759s.

Masubuchi, K. (1980). Weld cracking and joint restraint. In: *Analysis of Welded Structures*, 518–576. Oxford, England: Pergamon Press.

McLean, D. (1962). *Mechanical Properties of Metals*. New York: Wiley.

Petch, H.J. (1952). Delayed fracture of metals under static load. *Nature* 169: 84–843.

Savage, W.F., Nippes, E.F., and Szekeres, E.S. (1976). Hydrogen induced cold cracking in a low alloy steel. *Welding Journal* 55: 276s–283s.

Shewmon, P.G. (1983). *Diffusion in Solids*. New York: McGraw-Hill.

Trivedi, R. and Somboonsuk, K. (1984). Constrained dendritic growth and spacing. *Materials Science and Engineering* 65: 65–74.

Troiano, A.R. (1960). The role of hydrogen and other interstitials in the mechanical behavior of metals. *Trans ASM* 52: 54–80.

van Vlack, L.H. (1975). *Elements of Materials Science and Engineering*. Reading, MA: Addison-Wesley.

Voldrich, C.B. (1947). Cold cracking in the heat-affected zone. *Welding Journal* 26: 152s–169s.

6 Solidification

Solidification is important in laser processing, since the behavior of the molten pool during solidification determines the fusion zone grain shape and size, microstructure, defects, and ultimately, the mechanical properties of the part. In Chapter 5, we briefly outlined the basic concepts of solidification, focusing on nucleation. The criterion that determines which structure prevails in the solidified material was also introduced. Our focus in this chapter will be on growth, on the assumption that nucleation has already taken place.

For processes where local equilibrium is maintained at the solid–liquid interface, conventional solidification theory is directly applicable. However, for those processes that result in rapid solidification, such theories may not be directly applicable. For completeness, we shall discuss both types of solidification systems, starting with conventional solidification processes. Under conventional solidification, we start with the simpler case of solidification without flow, where we first address solidification of a pure metal, followed by that of a binary alloy. As part of our discussion on binary alloy solidification, we further discuss the conditions that lead to the formation of different types of microstructures (planar, cellular, and dendritic), and also the mushy zone (where solid and liquid coexist). Then we provide some insight on the basic concepts of solidification with flow. Finally, we outline the specific issues associated with rapid solidification.

6.1 SOLIDIFICATION WITHOUT FLOW

6.1.1 Solidification of a Pure Metal

We start the discussion by considering the simplest case, the classical Stefan problem. This involves a one-dimensional analysis of the solidification of either a pure metal or binary melt of eutectic composition. Let us consider a semi-infinite liquid material of initial uniform temperature, T_0. Now let the base of the material be suddenly placed in contact with a chilled surface of constant temperature, T_c at time $t = 0$, such that $T_c < T_m$, where T_m is the melting temperature of the liquid material. We set up a coordinate system as shown in Figure 6.1 with origin at the chilled surface, $x = 0$. Since we are considering a pure metal or eutectic, the composition of the solid that forms and that of the liquid are identical, so composition does not influence solidification. We further assume that there is no flow in the liquid and neglect any density differences that might exist between the solid and liquid phases.

The problem is one of determining the temperature field, $T(x, t)$, in the solid and liquid phases, and also the motion of the solid–liquid interface, $s(t)$. Since this is a pure metal, the interface is a

Principles of Laser Materials Processing: Developments and Applications, Second Edition. Elijah Kannatey-Asibu, Jr.
© 2023 John Wiley & Sons, Inc. Published 2023 by John Wiley & Sons, Inc.
Companion Website: www.wiley.com/go/Kannatey-Asibu/PrinicplesLaserMaterialsProcessing

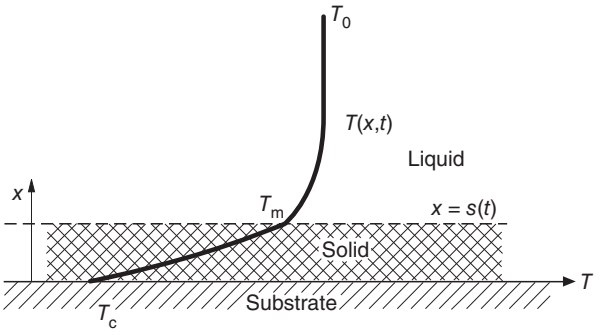


Figure 6.1 Temperature field for one-dimensional solidification of a pure material. Source: From Huppert (1990)/with permission of Cambridge University Press.

distinct front that separates the solid and liquid phases. Determination of the temperature field and interface requires solving the one-dimensional conduction equation in the two phases:

$$\rho_s c_{ps} \frac{dT}{dt} = \frac{\partial}{\partial x}\left(k_s \frac{\partial T}{\partial x}\right) \quad (0 < x \leq s(t)) \tag{6.1a}$$

$$\rho_l c_{pl} \frac{dT}{dt} = \frac{\partial}{\partial x}\left(k_l \frac{\partial T}{\partial x}\right) \quad (x \geq s(t)) \tag{6.1b}$$

with the initial and boundary conditions:

$$T = T_0 \, (t = 0) \tag{6.2a}$$

$$T = T_c \, (x = 0) \tag{6.2b}$$

$$T \to T_0 \quad (x \to \infty) \tag{6.2c}$$

where the subscripts s and l indicate solid and liquid, respectively, and T_c is the temperature of the chilled surface.

Another boundary condition is given by the need to have conservation of heat flux (energy balance) at the interface between the liquid and solid phases:

$$L_{mv} u_x = k_s G_{Ts} - k_l G_{Tl} \tag{6.2d}$$

where $G_{Ts} = \left.\frac{dT}{dx}\right|_{s-}$ is temperature gradient in the solid at the interface, $G_{Tl} = \left.\frac{dT}{dx}\right|_{s+}$ is temperature gradient in the liquid at the interface, L_{mv} is the latent heat of solidification (melting) per unit volume of solid, and $u_x = \dot{s} = SR$ is the interface velocity.

Equation (6.2d) implies that the net difference between the heat fluxes in the solid and liquid phases at the interface is the rate of release of the latent heat of solidification. Very few analytical solutions exist for Eqs. (6.1a–6.2b), and they are normally solved using numerical methods. The temperature field that results from their solution is of the form as shown in Figure 6.1.

The discussion thus far has focused on a pure material where solidification occurs at constant temperature, and the evolution of latent heat occurs as a step change. Most materials used in practice, however, are alloys. The solidification process for such materials takes place over a temperature range, between the liquidus and solidus (or eutectic) temperatures. The latent heat that evolves then has a functional relationship with temperature. In the next section, we discuss the solidification of alloys, focusing on the simpler but yet very common type, a binary alloy.

6.1.2 Solidification of a Binary Alloy

In this section, we first discuss the simple case where the solidification interface is planar. This will be followed by a detailed discussion of the conditions that lead to interface instability. Finally, the mushy zone that exists between the completely solid and completely liquid interfaces will be discussed.

6.1.2.1 Temperature and Concentration Variation in a Solidifying Alloy Let us consider a setup identical to that of the pure metal solidification, except that the material is now an alloy of two elements. We make the following assumptions:

1. The composition of the solid does not change.
2. The solid–liquid interface is flat and distinct, that is, there is no mushy zone.
3. There is no convection in the liquid.
4. Transport occurs only by molecular diffusion and is one-dimensional.
5. Local thermodynamic equilibrium exists. That means the concentration of solute in the solid and liquid sides of the interface is given by the equilibrium phase diagram.

The governing heat conduction equations are the same as for a pure metal, Eqs. (6.1a) and (6.1b), with the thermal boundary conditions, Eqs. (6.2a–6.2c), and the thermal conservation condition, Eq. (6.2d).

As solidification proceeds, solute is rejected from the interface into the liquid. Thus, in addition to the energy Eqs. (6.1a and 6.1b), we must also have a solute transport equation that describes the solute concentration in the liquid and is given by:

$$\frac{dC}{dt} = \frac{\partial}{\partial x}\left(D_l \frac{\partial C}{\partial x}\right) \quad (x \geq s(t)) \tag{6.3}$$

where C is the solute concentration (%) and D_l is the diffusion coefficient of the solute in the liquid, (m^2/s).

Considering conservation of solute at the interface, we have the following boundary condition:

$$u_x C + D_l \left.\frac{dC}{dx}\right|_{s+} = 0 \quad (x = s(t)+) \tag{6.4}$$

The other boundary and initial conditions are

$$C \to C_0 \quad (x \to \infty)$$

$$C = C_0 \, (t = 0)$$

where C_0 is the initial liquid composition.

Since solidification occurs under local thermodynamic equilibrium, the solute concentration at the interface and solidification or melting temperature are coupled by a simplified relationship of the form:

$$T = -m_g C + b \tag{6.5}$$

where m_g is the absolute value of the slope of the liquidus curve, which is approximated as a straight line.

Numerical solution of Eqs. (6.3–6.5) along with (6.1a–6.2b) gives temperature and concentration profiles of the form as shown in Figure 6.2.

For the ideal situation described in the preceding paragraphs, the temperature gradient is in the direction of crystal growth. The solid–liquid interface is thus planar and moves with a steady

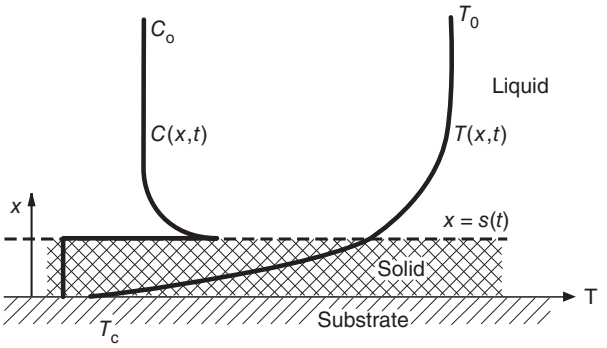


Figure 6.2 Schematic temperature field for one-dimensional solidification of a binary alloy. Source: From Huppert (1990)/with permission of Cambridge University Press.

velocity, u_x. The process is then one-dimensional, and the solute and temperature fields vary only in the growth direction. If the liquid state is quiescent (that is, at rest), the concentration and temperature extending ahead of the interface are exponential.

As briefly discussed in Section 5.1.1.2, the interface that develops during solidification of noneutectic alloy compositions is not always flat (planar) as assumed but may become cellular or dendritic as a result of instabilities associated with perturbations at the interface.

6.1.2.2 Interface Stability Theories There are two classes of theory that are used in analyzing *interface instability*, or the conditions under which deviations from the planar interface will occur. These are

1. The constitutional supercooling theory.
2. The morphological stability theory.

The basic principles of each approach are outlined in the following sections.

Constitutional Supercooling Theory This is normally used for solidification processes that occur at low interface velocities and is based on thermodynamics principles. The rationale is to determine whether the liquid ahead of a moving interface is undercooled with respect to its composition. Such undercooling may result from rejection of solute by the growing solid and tends to make the interface unstable, resulting in the formation of cells or dendrites that protrude into the liquid. The planar interface may be shown to be stable if

$$k_s G_{Ts} + k_l G_{Tl} > m_g G_c (k_s + k_l) f(A, k_p) \tag{6.6}$$

where $A = \frac{k_p^2 T_m \gamma_e u_x}{m_g (k_p - 1) C_0 D_l L_{mv}}$, $f(A, k_p) \approx 1 - A, \quad A \leq 1, G_c = \frac{u_x C_0 (k_p - 1)}{D_l k_p}$ is the solute concentration gradient at the interface in the unperturbed liquid, neglecting diffusion in the solid, k_p is the equilibrium partition coefficient, T_m is the interface melting temperature in the absence of solute and curvature, and γ_e is the interface surface energy.

Typically, solidification velocities are such that $A \ll 1$, in which case $f(A, k_p) \approx 1$. The stability criterion then reduces to the form:

$$k_s G_{Ts} + k_l G_{Tl} > m_g G_c (k_s + k_l) \text{ or } k_s G_{Ts} + k_l G_{Tl} > m_g \frac{u_x C_0 (k_p - 1)}{D_l k_p} (k_s + k_l) \tag{6.7}$$

which can be simplified by assuming $G_{Ts} \approx G_{Tl}$, giving

$$\frac{G_{Tl}}{u_x} > \frac{m_g C_0}{D_l} \frac{(k_p - 1)}{k_p} \quad \text{or} \quad G_{Tl} > \frac{m_g u_x C_0}{D_l} \frac{(k_p - 1)}{k_p} \tag{6.8}$$

On the contrary, at high solidification velocities $f(A, k_p) = 0$, since A cannot exceed 1. The stability criterion then becomes

$$k_s G_{Ts} + k_l G_{Tl} > 0 \tag{6.9}$$

The constitutional supercooling criterion, represented by Eqs. (6.7) or (6.8), is illustrated in Figure 5.7 as the line between the planar and cellular regions. In the planar region, the actual temperature gradient in the liquid is greater than the equilibrium temperature gradient as obtained from the right-hand side of Eq. (6.8), and thus planar growth is morphologically stable. In the cellular region, the solute-gradient effect dominates, and the actual temperature gradient is less than that given by the right-hand side of Eq. (6.8). This results in chance protuberances becoming pronounced, leading to cellular growth, which results in microsegregation. At yet higher interface velocities relative to the temperature gradient, or higher solute concentrations, growth becomes dendritic.

The constitutional supercooling theory is a limiting case of the more general theory of morphological stability and is adequate for a number of conventional solidification processes where local equilibrium conditions hold, and the solidification rates are relatively low. However, since it neglects surface tension effects and latent heat evolution, both of which have a stabilizing effect on the interface, it is generally not adequate for modeling rapid solidification processes. In fact, Eq. (6.7) predicts a less stable interface as the interface velocity increases, which contradicts experimental results that show stable flat interface morphologies at high interface velocities. For such cases, the morphological stability theory, which incorporates surface energy effects, is more appropriate.

Morphological Stability Theory. The morphological stability and constitutional supercooling theories predict the same outcome for a significant number of conventional solidification processes. However, the morphological stability theory predicts a more stable interface during rapid solidification due to the stabilizing influence of surface tension and latent heat evolution.

Let us assume a perturbation at the solidification front of the form:

$$x = \delta e^{(\omega_t t + i\omega_y y + i\omega_z z)} \tag{6.10}$$

where δ is the amplitude of perturbation at time $t = 0$, ω_t is the temporal frequency, and ω_y and ω_z are spatial frequencies.

Using a linear time-dependent stability analysis, it can be shown that the interface is stable when the real part of ω_t is negative for all perturbations. The value of ω_t can be obtained by solving the steady-state heat flow and diffusion equations with appropriate boundary conditions. In other words, the interface is stable with respect to infinitesimal perturbations when

$$m_g G_c \frac{\omega^* - \frac{u_x}{D_l}}{\omega^* - \frac{(1-k_p)u_x}{D_l}} - \frac{T_m \gamma_e \omega^2}{L_{mv}} - \left(\frac{k_s G_{Ts} + k_l G_{Tl}}{k_s + k_l} \right) < 0 \tag{6.11}$$

where $\omega^* = \frac{u_x}{2D_l} + \left[\left(\frac{u_x}{2D_l} \right)^2 + \omega^2 \right]^{1/2}$, and $\omega^2 = \omega_y^2 + \omega_z^2$

The first term of Eq. (6.11) is due to the solute concentration, the second due to capillary effects, and the third due to thermal effects. The interface is stable when the inequality is satisfied. The solute term thus tends to make the interface unstable, while the capillary term tends to prevent the growth of a perturbation, and since the thermal term may be either positive or negative, it can be either stabilizing or destabilizing, depending on whether the numerator is negative or positive, that is, if G_{Ts} is sufficiently positive.

It can be shown that the morphological stability criterion, Eq. (6.11), reduces to the constitutional supercooling criterion, Eq. (6.7), in the limiting case when the interface velocity is very low, and when capillary effects are neglected. The relationship between the constitutional and morphological stability criteria is illustrated in Figure 6.3. It shows the critical concentration of copper below which interface stability occurs for different interface velocities in directional solidification of an Al–Cu alloy.

At high interface velocities, when $\omega \ll \frac{u_x}{D_{\mathrm{l}}}$, and for a stabilizing thermal field, the morphological stability criterion can be shown to reduce to

$$m_{\mathrm{g}} G_{\mathrm{c}} \left(\frac{{D_{\mathrm{l}}}^2}{{u_x}^2 k_{\mathrm{p}}} \right) \omega^2 - \frac{T_{\mathrm{m}} \gamma_{\mathrm{e}} \omega^2}{L_{\mathrm{mv}}} < 0 \tag{6.12}$$

for a stable interface. This is referred to as the absolute stability criterion (Figure 6.3) and is essentially independent of G_{Tl}.

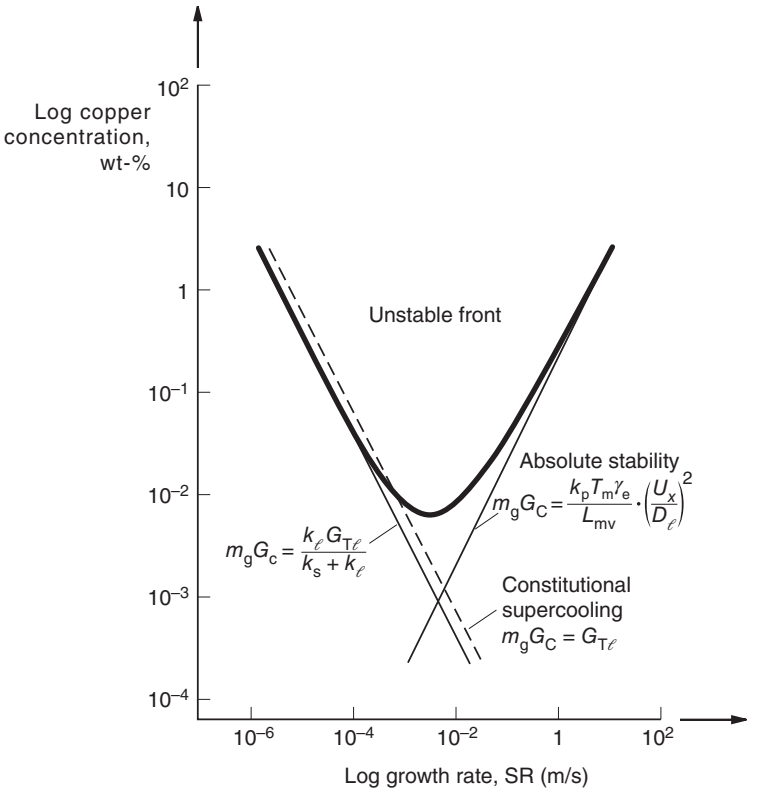


Figure 6.3 Relationship between morphological and constitutional supercooling stability criteria for Al–Cu alloys ($G_{\mathrm{Tl}} = 2 \times 10^4$ K/m). Source: From Mehrabian (1982)/with permission of Taylor & Francis.

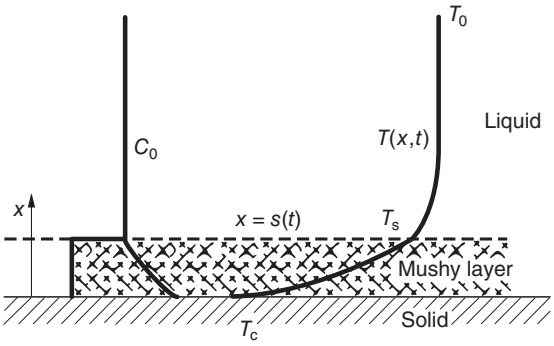


Figure 6.4 Schematic temperature field for one-dimensional solidification of a binary alloy with a mushy zone. Source: From Huppert (1990)/with permission of Cambridge University Press.

6.1.2.3 Mushy Zone As opposed to the solidification of pure materials, binary alloys do not exhibit a distinct front separating the solid and liquid phases. During solidification of actual alloys, three regions exist in the overall solidification domain. These include the solid and liquid phases and in between them is the mushy zone. The latter is a region of finite size where solid and liquid coexist and is bounded by the liquidus and solidus (or eutectic) curves. For simplicity, we assume that the mushy zone exists between the interface and the base surface (Figure 6.4). We further assume that the volume fraction, f_s, of the solid in the mushy zone is constant. The governing equations for the temperature field are then given by:

$$\overline{\rho c}_p \frac{dT}{dt} = \frac{\partial}{\partial x}\left(\overline{k}\frac{\partial T}{\partial x}\right) \quad (0 < x \le s(t)) \tag{6.13}$$

$$\rho_l c_{pl} \frac{dT}{dt} = \frac{\partial}{\partial x}\left(k_l \frac{\partial T}{\partial x}\right) \quad (x \ge s(t)) \tag{6.14}$$

$$L_{mv} f_s u_x = \overline{k}_s G_{Ts} - k_l G_{Tl} \tag{6.15}$$

where the overbar indicates properties in the mushy region and are based on averages weighted by the volume fraction. For example,

$$\overline{k} = f_s k_s + f_l k_l \tag{6.16}$$

and is an appropriate approximation for a random mixture of solid and liquid. Here, f stands for volume fraction.

The solid in the mushy region is assumed to be of constant concentration, while the interstitial liquid concentration corresponds to that of the liquidus temperature:

$$T = -m_g C + b \quad (0 < x \le s(t)) \tag{6.17}$$

The concentration of the liquid phase is also assumed to be equal to its original concentration, since diffusion effects are considered negligible:

$$C = C_0 \quad (x \ge s(t)) \tag{6.18}$$

For global conservation of solute (i.e. the total amount of solute must be conserved), we have

$$(1 - f_s)\int_0^{s(t)} C(x,t)dx = s(t)C_0 \tag{6.19}$$

The temperature and concentration profiles for this case are shown in Figure 6.4.

When the volume fraction is variable, the governing equations in the mushy zone become

$$\overline{\rho c}_{\mathrm{p}} \frac{\partial T}{\partial t} = \frac{\partial}{\partial x}\left(\bar{k}\frac{\partial T}{\partial x}\right) + L_{\mathrm{mv}}\frac{\partial f_{\mathrm{s}}}{\partial t} \tag{6.20}$$

$$(1 - f_{\mathrm{s}})\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\left(D_{\mathrm{l}}(1 - f_{\mathrm{s}})\frac{\partial C}{\partial x}\right) + C\frac{\partial f_{\mathrm{s}}}{\partial t} \tag{6.21}$$

The second term on the right-hand side of Eq. (6.20) represents latent heat release into the mushy zone, while that of Eq. (6.21) represents the release of solute into the interstitial fluid. The boundary condition for the spatial derivative of the solid volume fraction is given by:

$$\frac{\partial T}{\partial x} = -m_{\mathrm{g}}\frac{\partial C}{\partial x} \quad (x = s(t)+) \tag{6.22}$$

which corresponds to equilibrium saturation in the liquid just ahead of the growing dendrites.

Since there are two phases, a natural approach to modeling such a system is to develop separate governing equations for each phase (solid and liquid in this case) and then couple the two sets of equations using boundary conditions at the interface. This is the approach that has thus far been used in this section and is commonly referred to as the multiple domain method. Incorporation of the volume fraction of each phase into the governing equations enables the mushy zone to be analyzed. However, the interfacial geometry between the solid and liquid phases is quite complex, and it is also an unknown function of space and time for which a solution needs to be obtained. The two-phase approach is thus very difficult to use for all but the simplest configurations such as pure materials. To circumvent this problem and simplify the analysis, one-phase or continuum models can be developed from the two-phase governing equations, resulting in the entire domain being treated as a single region governed by one set of conservation equations, which are valid for the liquid and solid regions, as well as the mushy zone. Thus, with the continuum models, it is not necessary to track the interface, and there is no need for boundary conditions internal to the solidifying medium. They are therefore suitable for analyzing alloys. The formulation for the continuum models is based on volume-averaging techniques where the microscopic equations, which are valid for each phase, are integrated over a small volume element. Classical mixture theory is also used.

6.2 SOLIDIFICATION WITH FLOW

The discussion in the preceding section focused on solidification in a liquid medium where no flow occurs. In a number of applications such as cladding, significant flow occurs in the liquid medium as it solidifies. Such flow may be due to either natural convection (from the solidification process itself) or forced convection (induced by external factors). In either case, the resulting motion in the melt does affect the solidification process. Natural convection results from buoyancy forces that are caused by density variations within the fluid. Such variations in density may be induced by either temperature or compositional (solutal) gradients that occur during the process. Compositional gradients result from solute rejection into the liquid. Thus, we can have thermal or solutal buoyancy, or both.

In such situations where flow occurs in the solidification medium, the governing equations are

1. Mass.
2. Momentum.
3. Energy.
4. Solute concentration equations.

For simplicity, we make the following assumptions:

1. Flow is Newtonian and laminar.
2. The properties of each phase are homogeneous and isotropic.
3. Local thermodynamic equilibrium exists at the solid–liquid interface, that is, at the interface, $T = T_s = T_l$ and $C_s = k_p\, C_l$.
4. Negligible field disturbances such as dispersion fluxes and surpercooling of the liquid.
5. Saturated mixture conditions exist, that is, $f_s + f_l = 1$. That means there cannot be void formation. Void formation occurs when $\rho_s > \rho_l$.
6. Solid phase is free of internal stresses and does not deform. Internal stresses develop when $\rho_s < \rho_l$.
7. Negligible solute diffusion in the solid.

We also note that there are two principal driving forces that influence flow in the molten pool during laser processing, and these are as follows:

1. Buoyancy forces, which are accounted for in the momentum equations to be discussed shortly.
2. Shear and pressure that act on the pool surface as a result of surface tension gradient and surface curvature. These are modeled as boundary conditions of the governing equations.

Since three-dimensional analysis can be computationally intensive, we simplify things by only considering two-dimensional analysis. The governing equations for the one-phase or continuum model are then given by:

Conservation of mass: The general form of the governing equation for mass conservation is

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0 \tag{6.23}$$

where $\mathbf{u}$ is the velocity vector.

Momentum equations: The momentum equations are modified by a concentration term, and for two-dimensional analysis, they become

$$\frac{\partial}{\partial t}(\rho u_x) + \nabla \cdot (\rho \mathbf{u} u_x) = -\frac{\partial P}{\partial x} + \nabla \cdot (\mu \nabla u_x) + F_x \tag{6.24a}$$

$$\frac{\partial}{\partial t}(\rho u_z) + \nabla \cdot (\rho \mathbf{u} u_z) = -\frac{\partial P}{\partial z} + \nabla \cdot (\mu \nabla u_z) + F_z - \rho g[\beta_T (T - T_0) + \beta_c (C_l - C_0)] \tag{6.24b}$$

where β_T is the thermal volume expansion coefficient (/K), β_c is the solutal volume expansion coefficient (/%), g is the acceleration due to gravity (m/s^2), P is the pressure (Pa (N/m^2)), F_x and F_z are the external body forces in the x and z directions, respectively (N/m^3), μ is the viscosity (Pa s), u_x and u_z are the magnitudes of velocities in the x and z directions, respectively (m/s), and T_0 and C_0 are the reference (usually eutectic) temperature and concentration, respectively.

The last term on the right-hand side of Eq. (6.24b) is the force due to thermal and solutal buoyancy.

Enthalpy (*energy*) *conservation*: For convenience, the energy equation is expressed in terms of enthalpy, h_e:

$$\frac{\partial}{\partial t}(\rho h_e) + \nabla \cdot (\rho \mathbf{u} h_e) = \nabla \cdot \left(\frac{k}{c_{ps}} \nabla h_e \right) + S_h \tag{6.25}$$

Solute concentration: The equation for solute concentration is expressed as:

$$\frac{\partial}{\partial t}(\rho C) + \nabla \cdot (\rho \mathbf{u} C) = \nabla \cdot (D \nabla C) + S_c \tag{6.26}$$

where S_h and S_c are enthalpy and concentration source terms, respectively.

In this formulation, the enthalpy, h_e, is defined as:

$$h_e = \int_0^T c_{ps} dT \tag{6.27a}$$

Thus,

$$h_s = h_e \quad \text{and} \quad h_l = h_e + \delta h_e \tag{6.27b}$$

$$\delta h_e = \int_0^T (c_{pl} - c_{ps}) dT + L_m \tag{6.27c}$$

where c_{pl} and c_{ps} are the specific heats of the liquid and solid, respectively (J/kg K), h_l and h_s are the enthalpies of the liquid and solid, respectively (J/kg), and L_m is the latent heat of fusion per unit mass (J/kg).

The exact form of the one-phase model that is used depends on the nature of the mushy zone. Thus, at this point, we consider two limiting cases:

1. Mushy fluid.
2. Columnar dendritic structure.

6.2.1 Mushy Fluid

This type of mushy zone is similar to granular solid particles that are dispersed within a liquid medium (Figure 6.5). It is more commonly found in the equiaxed zone of some metals and also in amorphous materials such as glasses and waxes. Both the solid and liquid phases then have the same velocity, and as such, this model is sometimes referred to as the no relative motion model:

$$\mathbf{u} = \mathbf{u}_l = \mathbf{u}_s \tag{6.24}$$

where $\mathbf{u}_l$ and $\mathbf{u}_s$ are the liquid and solid velocities, respectively.

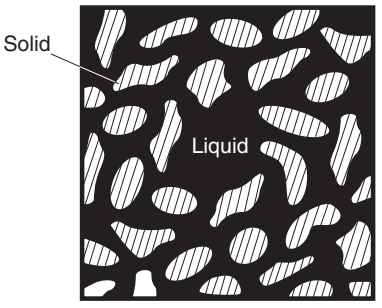


Figure 6.5 Mushy zone with solid particles dispersed within the liquid medium.

Since the volume fraction of the solid varies in the mushy zone, an effective viscosity, $\overline{\mu}$, is used at any location as:

$$\overline{\mu} = f_s \mu_s + f_l \mu_l \tag{6.25}$$

where f_s and f_l are the solid and liquid volume fractions, respectively, and are related to the respective mass fractions, m_s and m_l, by:

$$\overline{\rho} m_s = \rho_s f_s \quad \text{and} \quad \overline{\rho} m_l = \rho_l f_l \tag{6.29a}$$

where ρ_l and ρ_s are the liquid and solid densities, respectively, and $\overline{\rho}$ is the overall mixture density and is given by:

$$\overline{\rho} = \rho_s f_s + \rho_l f_l \tag{6.29b}$$

Since there is no flow in the fully solid region, the velocity there has to be zero. In other words, the system velocity has to approach zero as the local liquid fraction approaches zero. This is normally achieved by specifying a very large solid viscosity, μ_s, say of the order of 10^8, such that as the solid volume fraction tends to unity, the effective viscosity $\overline{\mu}$ dominates the momentum equations and eventually reduces the velocity $\mathbf{u}$ to zero when fully solidified. The momentum equations for this case are given by:

$$\frac{\partial}{\partial t}(\overline{\rho} u_x) + \nabla \cdot (\overline{\rho} \mathbf{u} u_x) = -\frac{\partial P}{\partial x} + \nabla \cdot (\overline{\mu} \nabla u_x) \tag{6.30a}$$

$$\frac{\partial}{\partial t}(\overline{\rho} u_z) + \nabla \cdot (\overline{\rho} \mathbf{u} u_z) = -\frac{\partial P}{\partial z} + \nabla \cdot (\overline{\mu} \nabla u_z) - \overline{\rho} g [\beta_T (T - T_0) + \beta_c (C_l - C_0)] \tag{6.30b}$$

The corresponding energy or enthalpy equation is given by:

$$\frac{\partial}{\partial t}(\overline{\rho} h_e) + \nabla \cdot (\overline{\rho} \mathbf{u} h_e) = \nabla \cdot \left(\frac{k}{c_{ps}} \nabla h_e \right) + \nabla \cdot \left(\frac{k}{c_{ps}} \nabla (h_s - h_e) \right) \tag{6.31}$$

The terms on the right-hand side of the equation represent the net Fourier heat flux. The solute concentration is given by:

$$\frac{\partial}{\partial t}(\overline{\rho} C) + \nabla \cdot (\overline{\rho} \mathbf{u} C) = \nabla \cdot (\overline{\rho} D \nabla C) + \nabla \cdot (\overline{\rho} D \nabla (C_l - C)) \tag{6.32}$$

6.2.2 Columnar Dendritic Structure

The columnar dendritic structure is such that the solid and liquid phases are distinct (Figure 6.6), with the solid having a specified velocity which may or may not be zero, and is definitely different from the liquid velocity. The model that results from this formulation is sometimes referred to as the relative motion model. Laser spot welding is an example of the case with zero solid velocity, while laser seam welding is for the case with nonzero solid velocity.

This structure is essentially similar to a porous medium, and as such, flow in the mushy zone is considered to be governed by Darcy's law. This effect is accounted for by modifying the momentum equations to account for flow in the porous medium and requires the following assumptions:

1. The solid is free of internal stresses.
2. Viscous stresses resulting from local density gradients are small.
3. The density in each phase is constant.

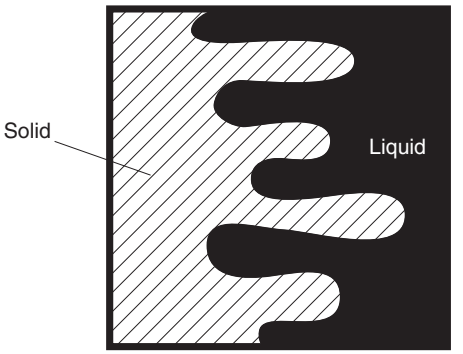


Figure 6.6 Columnar dendritic mushy zone at the solidifying interface.

With these assumptions, the momentum equations are given by:

$$\frac{\partial}{\partial t}(\rho u_x) + \nabla \cdot (\rho \mathbf{u} u_x) = -\frac{\partial P}{\partial x} + \nabla \cdot \left(\mu_l \frac{\rho}{\rho_l} \nabla u_x \right) - \frac{\mu_l}{K_p} \frac{\rho}{\rho_l} (u_x - u_{xs}) \tag{6.33a}$$

$$\frac{\partial}{\partial t}(\rho u_z) + \nabla \cdot (\rho \mathbf{u} u_z) = -\frac{\partial P}{\partial z} + \nabla \cdot \left(\mu_l \frac{\rho}{\rho_l} \nabla u_z \right) - \frac{\mu_l}{K_p} \frac{\rho}{\rho_l} (u_z - u_{zs}) - \rho g [\beta_T (T - T_0) + \beta_c (C_l - C_0)] \tag{6.33b}$$

where K_p represents permeability and is assumed to be isotropic. It depends on the cooling rate, temperature gradient, and primary dendrite arm spacing during the phase change. u_{is} represents solid velocity and is set to zero for static phase change systems, or to a constant value for continuous phase change systems.

There are a number of models available for the permeability, K_p. A common isotropic formulation is based on the Carman–Kozeny equation:

$$K_p = K_0 \left[\frac{f_l^3}{(1 - f_l)^2} \right] \tag{6.34}$$

where K_0 is a constant that depends on the mushy zone morphology. The Carman–Kozeny equation is generally valid for laminar flow and for liquid volume fractions less than 0.5.

The energy or enthalpy equation is also given by:

$$\frac{\partial}{\partial t}(\rho h_e) + \nabla \cdot (\rho \mathbf{u} h_e) = \nabla \cdot \left(\frac{k}{c_{ps}} \nabla h_e \right) + \nabla \cdot \left(\frac{k}{c_{ps}} \nabla (h_s - h_e) \right) - \nabla \cdot [\rho (h_l - h_e)(\mathbf{u} - \mathbf{u}_s)] \tag{6.35}$$

The first two terms on the right-hand side of the equation represent the net Fourier heat flux, while the last term represents the energy flux associated with relative phase motion.

The solute concentration equation is given by:

$$\frac{\partial}{\partial t}(\rho C) + \nabla \cdot (\rho \mathbf{u} C) = \nabla \cdot (\rho D \nabla C) + \nabla \cdot [\rho D \nabla (C_l - C)] - \nabla \cdot [\rho (C_l - C)(\mathbf{u} - \mathbf{u}_s)] \tag{6.36}$$

The first two terms on the right-hand side of the equation represent the net diffusive flux, while the last term represents the species flux associated with relative phase motion.

6.3 RAPID SOLIDIFICATION

Rapid solidification results when the rate at which the molten material cools is so high that its temperature falls significantly below its equilibrium freezing temperature, resulting in extensive undercooling before solidification begins. The process then occurs under nonequilibrium conditions with a high solidification velocity, and the interface temperature becomes a function of time and process parameters. The assumption of local thermodynamic equilibrium is then invalid.

Some of the benefits of rapid solidification include

1. Fine grain structure and solidification substructure (dendrite and cell spacing).
2. Reduced microsegregation.
3. Extended solubility of alloying elements.
4. Metastable crystalline and amorphous phases – rapid solidification is the basis for obtaining surface modification of materials using a laser beam.

There are two extreme heat flow conditions that are associated with rapid solidification. In one case, the molten material is superheated, and so the heat of fusion is withdrawn through the solid, and the interface velocity is controlled primarily by the rate of external heat extraction. This would be the case for surface melting and melt spinning, for example, where the interface velocity is limited by the rate of heat extraction to the substrate, and the material thickness.

In the other case, the molten material is supercooled and thus acts as an effective heat sink, quickly removing any released latent heat. Crystal growth kinetics then become the main limiting factor. However, the latent heat that is released at the interface as solidification proceeds may result in reheating of the molten material to a temperature close to the equilibrium solidification temperature, thereby reducing the growth rate of the freezing front (interface velocity). This is referred to as recalescence and may result in microsegregation, reduced supersaturation, and phase segregation. The interface velocity for this case is determined primarily by the extent of supercooling, and the heat loss to the surroundings does not significantly affect the solidification process in the initial stages but may become significant as solidification proceeds.

These are illustrated in the enthalpy-temperature diagram (Figure 6.7). Equilibrium solidification essentially follows the isothermal path. Here, the latent heat that is released is used to maintain the specimen's temperature at the equilibrium value until solidification is completed.

The conditions under which local thermodynamic equilibrium are valid may be approximated by considering the diffusive velocity, u_D, of the solute in the liquid, which may be defined as:

$$u_D = \frac{D_i}{l_a} \tag{6.37}$$

where l_a is the interatomic spacing and D_i is the coefficient of interdiffusion across the interface.

As the interface velocity approaches the diffusive velocity, the solidification front begins overtaking the solute atoms, at which point local thermodynamic equilibrium is no more valid at the interface. Complete solute trapping results as the interface velocity exceeds the diffusive velocity. Generally, as conditions depart from local equilibrium interface stability is enhanced.

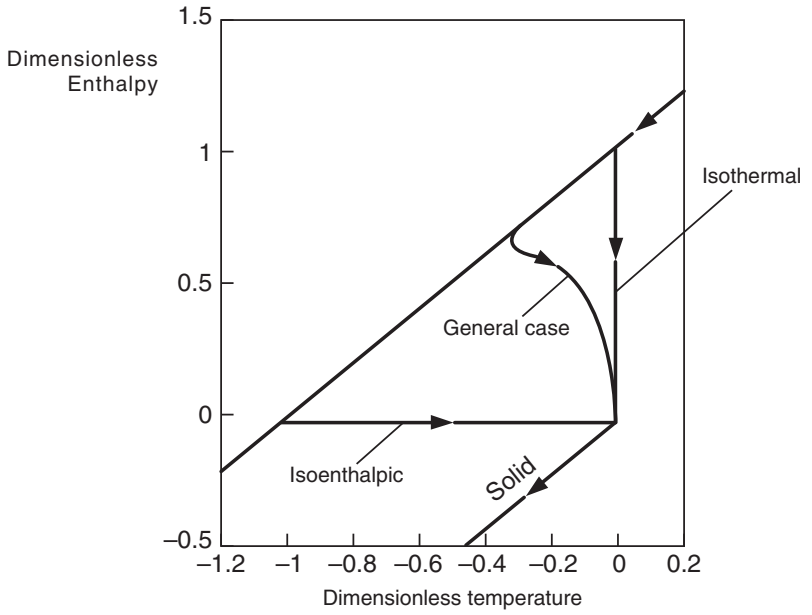


Figure 6.7 Enthalpy-temperature diagram. Source: From Mehrabian (1982)/with permission of Taylor & Francis.

6.4 SUMMARY

The emphasis in this chapter has been on the growth component of the solidification process, having covered the nucleation process in Chapter 5. For a pure metal without flow, the solidification interface is planar. That of an alloy may break down to form a cellular, columnar dendritic, or equiaxed dendritic structure as a result of instability associated with the process. The instability of the interface can be analyzed using either the constitutional supercooling theory or the morphological stability theory. Whereas the stability criterion based on the constitutional supercooling theory depends only on the temperature gradients in the liquid and solid at the interface, that of the morphological stability theory also takes into account the solute concentration as well as capillary effects. It is shown that the solute term tends to make the interface unstable, while the capillary term tends to prevent the growth of a perturbation, and since the thermal term may be either positive or negative, it can be either stabilizing or destabilizing. For analysis in the mushy zone, similar governing equations are used, except that they are modified to account for the solid volume fraction.

When solidification occurs under conditions involving flow, the analysis also includes mass and momentum conservation equations. The analysis is simplified by considering the mushy zone as either a mushy fluid (which is similar to granular solid particles that are dispersed within a liquid medium) or a columnar dendritic structure (where the solid and liquid phases are distinct, with the solid having a specified velocity which may or may not be zero, and is different from the liquid velocity).

With rapid solidification, which is the more likely scenario in laser processes that involve a large mass which acts as a heat sink, there are two extreme heat flow conditions:

1. One case where the molten material is superheated, which is also the more likely case for laser processing.
2. Another case where the molten material is supercooled.

Rapid solidification results in fine grain, dendrite and cell structures, reduced microsegregation, extended solubility of alloying elements, and metastable crystalline and amorphous phases.

APPENDIX 6.A

List of symbols used in the chapter

Symbol	Parameter	Units
C	Alloy composition or solute concentration	%
C_l	Liquid composition or concentration at the interface	%
C_0	Initial (reference) liquid composition	%
C_s	Solid composition or concentration at the interface	%
D_i	Diffusion coefficient of solute across the interface	m^2/s
G_c	Solute concentration gradient at the solid–liquid interface	%/m
h_e	Enthalpy	J
k_p	Equilibrium partition coefficient (see Appendix 6B)	—
K_p	Permeability	m^2
l	Subscript for liquid	—
l_a	Interatomic spacing	m
m_g	Absolute value of slope of liquidus curve	°C
s	Subscript for solid	—
$s(t)$	Position of solid–liquid interface	m
$\dot{s}$	Solidification rate or interface velocity	m/s
S_c	Concentration source term	$kg/m^3 s$
S_h	Enthalpy source term	W kg/m^3
T_c	Constant temperature of a chilled surface	°C
u_D	Diffusive velocity of solute in liquid	m/s
δ	Amplitude of perturbation at time $t = 0$	m
ω_t	Temporal frequency	Hz
$\omega_{y,z}$	Spatial frequencies	Hz

PROBLEMS

6.1. Show that the morphological stability criterion, Eq. (6.11), reduces to the constitutional supercooling criterion, Eq. (6.8), in the limiting case when the interface velocity is very low and capillary terms are neglected.

6.2. Show that at high interface velocities, and for a stabilizing thermal field, the morphological stability criterion reduces to

$$m_g G_c \left(\frac{{D_l}^2}{u_x^2 k_p} \right) \omega^2 - \frac{T_m \gamma_e \omega^2}{L_{mv}} < 0$$

6.3. Given

$$\frac{G_{Tl}}{SR} \leq \frac{(C_s - C_l)}{D_l} \frac{dT_l}{dC_l}$$

show that

$$\frac{G_{\mathrm{Tl}}}{SR} \leq C_0 \frac{(1-k_{\mathrm{p}})}{k_{\mathrm{p}} D_{\mathrm{l}}} \frac{dT_{\mathrm{l}}}{dC_{\mathrm{l}}}$$

6.4. Given that the concentration of solute atoms in the liquid in front of the solid–liquid interface during one-dimensional solidification in the x-direction is described by the following first-order differential equation:

$$D_{\mathrm{l}} \frac{dC_{\mathrm{l}}}{dx} = SR(C_{\mathrm{s}} - C_{\mathrm{l}})$$

show that the concentration profile is given by:

$$C_{\mathrm{l}} = C_0 \left[1 + \left(\frac{1}{k_{\mathrm{p}}} - 1 \right) e^{-\frac{SR}{D_{\mathrm{l}}} x} \right]$$

where C_0 is the original composition of the alloy in the liquid state, and the diffusion coefficient, D_{l}, is assumed constant.

6.5. A self-styled consultant claims that in modeling the behavior of the mushy zone, a reasonable approximation is to use the viscosity of the solid. For a 1.5% carbon steel, let the temperature of the mushy zone be 1325 °C.

(a) Investigate the consultant's claim and comment on its validity, using Eq. (6.29) and the lever rule.

(b) If the claim is judged to be invalid, suggest additional detail that could help support his/her suggestion.

BIBLIOGRAPHY

Beckermann, C. and Viskanta, R. (1988a). Double-diffusive convection during dendritic solidification of a binary mixture. *PCH PhysicoChemical Hydrodynamic* 10 (2): 195–213.

Beckermann, C. and Viskanta, R. (1988b). Natural convection solid/liquid phase change in porous media. *International Journal of Heat and Mass Transfer* 31 (1): 35–46.

Bennon, W.D. and Incropera, F.P. (1987a). A continuum model for momentum, heat and species transport in binary solid-liquid phase change systems - I. Model formulation. *International Journal of Heat and Mass Transfer* 30 (10): 2161–2170.

Bennon, W.D. and Incropera, F.P. (1987b). A continuum model for momentum, heat and species transport in binary solid-liquid phase change systems – II. Application to solidification in a rectangular cavity. *International Journal of Heat and Mass Transfer* 30 (10): 2171–2187.

Combeau, H., Roch, F., Poitrault, I. et al. (1990). Numerical study of heat and mass transfer during solidification of steel ingots. In: *Presented at Advanced Computational Methods in Heat Transfer*, 79–90. New York: Springer-Verlag.

Glicksman, M.E., Coriell, S.R., and McFadden, G.B. (1986). Interaction of flows with the crystal-melt interface. *Ann. Rev. Fluid Mechanics* 18: 307–335.

Huang, S.-C. and Glicksman, M.E. (1981). Fundamentals of dendritic solidification – I. *Steady-state tip growth, Acta Metallurgica* 29: 701–715.

Huppert, H.E. (1990). The fluid mechanics of solidification. *Journal of Fluid Mechanics* 212: 209–240.

Kou, S. (1987). *Welding Metallurgy*. New York: Wiley.

Kurz, W. and Fisher, D.J. (1992). *Fundamentals of Solidification*, 3e. Switzerland: Trans Tech Publications.

Mazumder, J. (1987). An overview of melt dynamics in laser processing. *High Power Lasers* 801: 228–241.

Mehrabian, R. (1982). Rapid solidification. *Metals Review* 27: 185–208.

Mullins, W.W. and Sekerka, R.F. (1964). Stability of a planar interface during solidification of a dilute binary alloy. *Journal of Applied Physics* 35: 444–451.

Oldenburg, C.M. and Spera, F.J. (1992). Hybrid model for solidification and convection. *Numerical Heat Transfer B* 21: 217–229.

Smith, T.J. and Hoadley, A.F.A. (1987). Recent developments in modeling metal flow and solidification. In: *Modelling the Flow and Solidification of Metals* (ed. T.J. Smith), 277–303. Dordrecht: Martinus Nijhoff Publishers.

Szekely, J. and Jassal, A.S. (1978). An experimental and analytical study of the solidification of a binary dendritic system. *Metallurgical Transactions B* 9b: 389–398.

Tiller, W.A., Jackson, K.A., Rutter, J.W., and Chalmers, B. (1953). The redistribution of solute atoms during the solidification of metals. *Acta Metallurgica* 1: 428–437.

Trivedi, R. and Somboonsuk, K. (1984). Constrained dendritic growth and spacing. *Materials Science and Engineering* 65: 65–74.

Tseng, A.A., Zou, J., Wang, H.P., and Hoole, S.R.H. (1992). Numerical modeling of macro and micro behaviors of materials in processing: a review. *Journal of Computational Physics* 102: 1–17.

Voller, V.R. and Prakash, C. (1987). A fixed grid numerical modeling methodology for convection-diffusion mushy region phase-change problems. *International Journal of Heat and Mass Transfer* 30 (8): 1709–1719.

Voller, V.R., Brent, A.D., and Prakash, C. (1989). The modelling of heat, mass and solute transport in solidificatiion systems. *International Journal of Heat and Mass Transfer* 32 (9): 1719–1731.

7 Residual Stresses and Distortion

Residual stresses are stresses that continue to exist in a material even when there are no external forces acting on it. They are produced in a number of ways, for example, during welding, grinding, forming, and heat treatment (especially when martensite is formed due to volumetric change, as austenite transforms to martensite). Residual stresses are important because they affect the strength of the part, either positively or negatively, and may also result in distortion.

Our discussion in this chapter starts with the fundamental causes of residual stresses, focusing on thermal stresses and nonuniform plastic deformation. This is followed by a basic stress analysis to explain the underlying mechanisms of residual stresses. The effects of residual stresses are then outlined, specifically strength effects and distortion. A discussion on techniques for measuring residual stresses (stress relaxation, X-ray diffraction, and neutron diffraction) then follows, and finally, methods for relieving residual stresses and distortion.

7.1 CAUSES OF RESIDUAL STRESSES

Two principal causes of residual stresses in materials processing are as follows:

1. Thermal stresses.
2. Nonuniform plastic deformation.

Our discussion will focus on these two causes. Other causes of residual stresses are nuclear radiation and electromagnetic field exposure.

7.1.1 Thermal Stresses

To illustrate the development of residual stresses as a result of temperature changes, let us consider three rods of equal length which are rigidly connected together at the top and bottom (Figure 7.1). Now, let the middle rod, 2, be heated uniformly to a high temperature and cooled. As it is heated, it tends to expand. If it were free, it would expand by an amount ΔL, given by:

$$\Delta L = \beta_l L_0 \Delta T \tag{7.1}$$

where L_0 is the original length of the rod, β_l is the linear coefficient of thermal expansion of the rod material (/K), and ΔT is the temperature rise in the rod.

Principles of Laser Materials Processing: Developments and Applications, Second Edition. Elijah Kannatey-Asibu, Jr.
© 2023 John Wiley & Sons, Inc. Published 2023 by John Wiley & Sons, Inc.
Companion Website: www.wiley.com/go/Kannatey-Asibu/PrinicplesLaserMaterialsProcessing

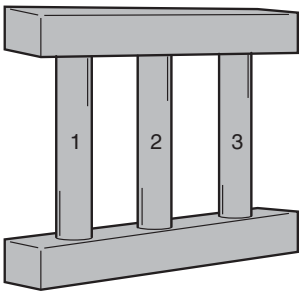


Figure 7.1 Three-rod frame with rods of equal length which are rigidly connected together at the top and bottom. Source: Jenney and O'Brien (2001)/with permission from American Welding Society.

However, it is prevented from doing so by the other two unheated rods. This results in compressive stress being induced in 2, while tensile stresses are induced in 1 and 3. Before the yield strength of the material is reached, the elastic strain (ε^{e}) induced in 2 is given by:

$$\varepsilon^{\mathrm{e}} = \frac{\Delta L}{L_0} = \beta_{\mathrm{l}} \Delta T \tag{7.2a}$$

which results in a compressive elastic stress of

$$\sigma = E_{\mathrm{m}} \beta_{\mathrm{l}} \Delta T \tag{7.2b}$$

where E_{m} is Young's modulus.

The stress magnitude in 2 keeps increasing with temperature until the yield stress in compression is reached (point Q) (Figure 7.2). At this time, plastic flow occurs. Beyond this point, the additional strain that is experienced, the plastic strain, ε^{p}, is given by:

$$\varepsilon^{\mathrm{p}} = \beta_{\mathrm{l}} \Delta T - \frac{S_{\mathrm{y}}}{E_{\mathrm{m}}} \tag{7.3}$$

where S_{y} is the yield strength of the material.

Since the yield strength decreases with increasing temperature, the stress in 2 decreases with further increase in temperature (QR).

On cooling, 2 tends to contract and again is restrained by 1 and 3. Since it is cooling, the yield strength increases as the temperature reduces further, and the stress in the rod becomes elastic. It quickly reduces in compressive stress and then becomes tensile, the tensile stress increasing elastically with further cooling until yielding (point S). Thus, at room temperature, rod 2 will be subjected to tensile residual stress, while 1 and 3 will be in compression. The residual stresses so produced are also called thermal stresses.

For a specific example in laser processing, we consider the development of residual stresses during welding. A schematic of a weld bead in progress along with temperature and stress distribution at various points along the bead are shown in Figure 7.3 at points A, B, C, and D.

Far ahead of the heat source at A–A, the temperature and stress distributions are zero. At section B–B, which passes through the heat source, very high temperatures exist in the immediate vicinity of the weld, and this tapers off to low temperatures away from the weld. Since the molten metal cannot endure stress, stresses in the weld are zero. Slightly farther off, there are low compressive stresses since the heated material close to the weld is prevented from expanding. Also, the yield stress is

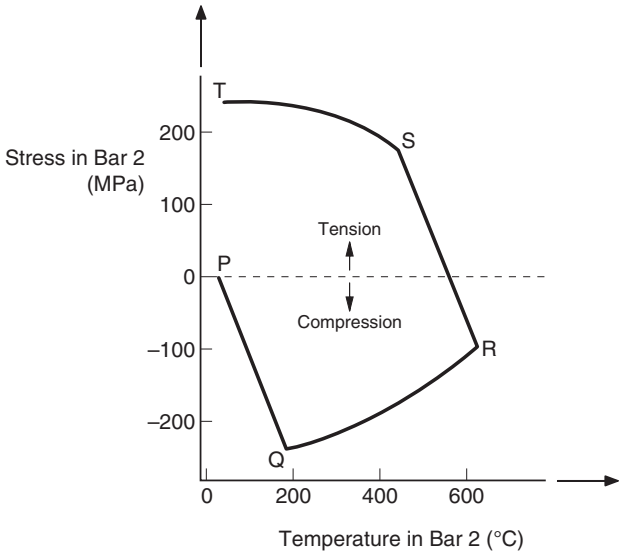


Figure 7.2 Stress–temperature curve for the middle rod of a three-rod frame. Source: Jenney and O'Brien (2001)/with permission from American Welding Society.

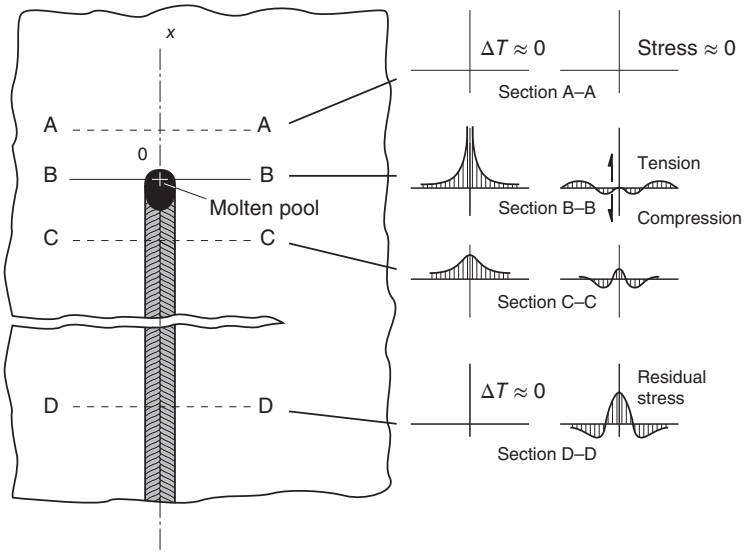


Figure 7.3 Stress and temperature changes during welding. Source: Jenney and O'Brien (2001)/with permission from American Welding Society.

low at the high temperatures and that makes the stresses low. Yet farther away, tensile stresses are developed that balance the compressive stresses close to the weld.

Behind the heat source, section C–C, lower temperatures exist in the weld bead, which then has a more uniform temperature distribution. The cooling weld bead and material close to it tends to contract, and this is prevented by the rest of the material. The result is tensile stresses in material within and close to the bead and balancing compressive stresses further away.

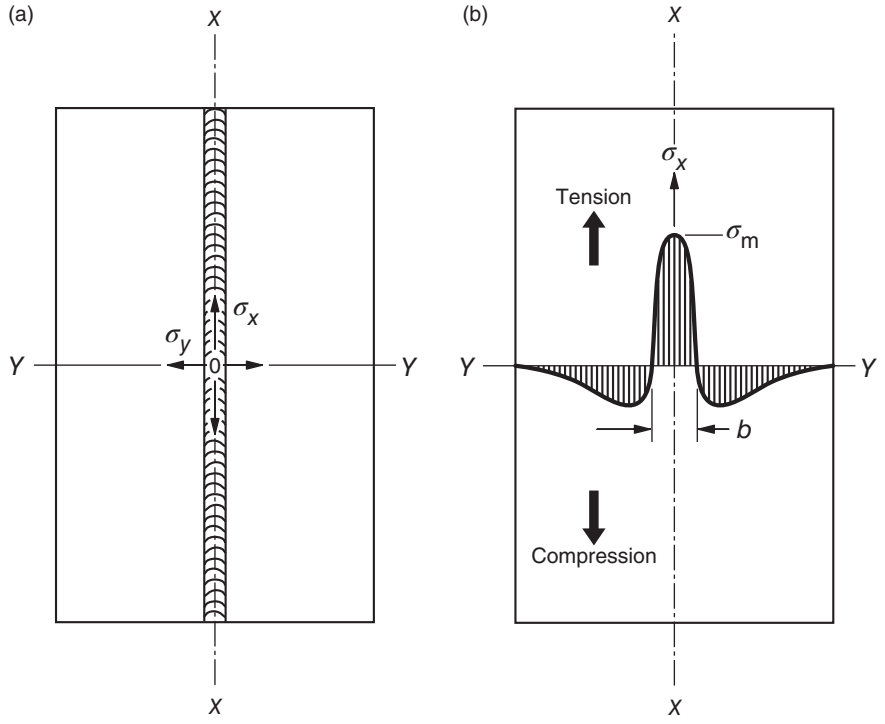


Figure 7.4 Longitudinal residual stress distribution in a weld. (a) Weld configuration. (b) Distribution of longitudinal stresses. Source: Jenney and O'Brien (2001)/with permission from American Welding Society.

Even further behind the heat source at section D–D, the temperatures are much lower, and the stresses become more pronounced. The longitudinal stress distribution that results at room temperature is shown in Figure 7.4. Important aspects of the longitudinal stress distribution (Figure 7.4b) are as follows:

1. Maximum tensile stress, σ_m.
2. Width of the tensile region, b.

7.1.2 Nonuniform Plastic Deformation

To understand how residual stresses are introduced as a result of plastic deformation, let us first consider a rod which is subjected to simple tension where the loading is high enough to produce uniform plastic deformation throughout the rod. If the applied load is subsequently released before necking occurs, there will be uniform elastic recovery throughout the rod. However, there will be permanent deformation within it after the load is removed (Figure 7.5), but no residual stresses will exist.

Now consider the beam shown in Figure 7.6a, which is subjected to bending. The bending moment, M_{init}, that will result in plastic deformation of the outer fibers is given by:

$$M_{init} = \frac{S_y b h^2}{6} \tag{7.4}$$

where b is the width and h is the thickness of the beam.

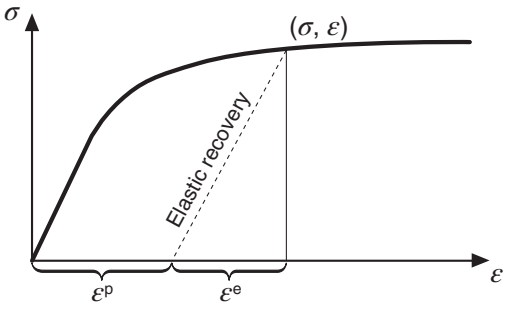


Figure 7.5 Stress–strain curve with elastic recovery and permanent set.

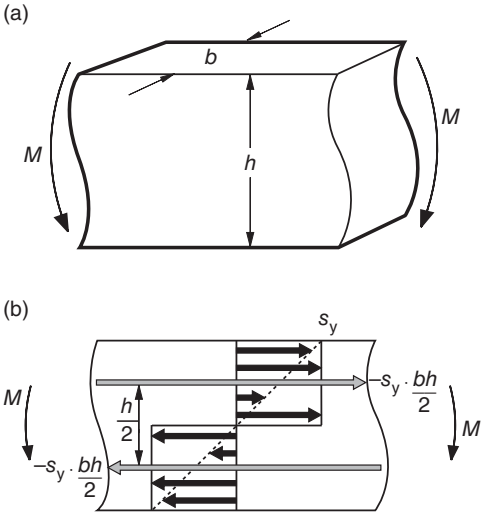


Figure 7.6 Beam bending and resulting stress distribution. (a) The beam. (b) The stress distribution.

The stress distribution in the beam is then given by the dashed line (Figure 7.6b). Assuming an elastic-perfectly plastic material, the stress distribution that results after full yield is given by the solid line (Figure 7.6b). The resulting force on either side of the beam (either the upper or lower half) is then given by $S_y bh/2$. Since the distance between the resulting forces is $h/2$, the bending moment M_y required to achieve full yield throughout the beam is

$$M_y = \frac{S_y bh}{2} \bullet \frac{h}{2} = \frac{S_y bh^2}{4} \tag{7.5}$$

If the external bending moment is now removed, elastic unloading occurs, and the stress distribution that results from elastic unloading is given by the dashed line in Figure 7.7a. The resultant stress distribution after unloading is obtained by superimposing the plastic stress distribution and the unloading curves, giving the final stress distribution within the beam after the load is removed, the residual stress (given by the solid line in Figure 7.7). The distribution in Figure 7.7a is the ideal case. The actual distribution is more like the one shown in Figure 7.7b.

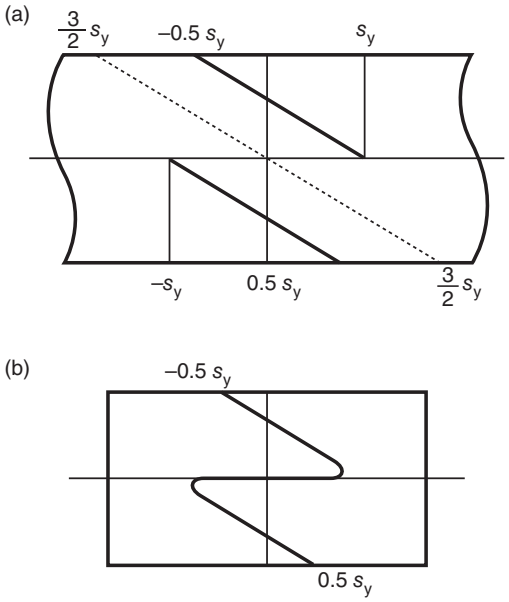


Figure 7.7 Stress distribution in a bent beam after unloading. (a) Ideal stress distribution. (b) Actual stress distribution.

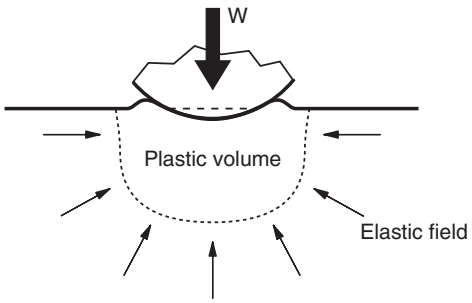


Figure 7.8 Indentation in a body.

Next, we consider a small region which is plastically deformed in a body by say, indentation (Figure 7.8). The immediate vicinity of the indentor will be subjected to plastic flow, and this will be surrounded by a large compressive elastic field. On removing the indentor, elastic unloading will occur, but there will be a permanent set (plastic strain that remains), which results in the dent left behind. The end result will be elastic compressive stresses that are left after unloading. Residual stresses due to nonuniform plastic deformation may thus result from hammering, rolling, dull cutting tools, and eroding particles.

7.2 BASIC STRESS ANALYSIS

In this section, we discuss some of the basic concepts necessary for analyzing residual stresses that may develop in materials.

7.2.1 Stress–Strain Relations

Due to space constraints, we shall limit our discussion in this section to isotropic materials and also to linear elastic behavior and plastic flow of metals.

7.2.1.1 Linear Elastic Behavior The stress–strain relations for linear elastic materials are given by Hooke's law:

$$\varepsilon_x = \frac{1}{E_{\mathrm{m}}}[\sigma_x - \nu_{\mathrm{p}}(\sigma_y + \sigma_z)] \tag{7.6a}$$

$$\varepsilon_y = \frac{1}{E_{\mathrm{m}}}[\sigma_y - \nu_{\mathrm{p}}(\sigma_z + \sigma_x)] \tag{7.6b}$$

$$\varepsilon_z = \frac{1}{E_{\mathrm{m}}}[\sigma_z - \nu_{\mathrm{p}}(\sigma_x + \sigma_y)] \tag{7.6c}$$

$$\gamma_{\mathrm{xy}} = \frac{\tau_{\mathrm{xy}}}{G} \tag{7.6d}$$

$$\gamma_{\mathrm{yz}} = \frac{\tau_{\mathrm{yz}}}{G} \tag{7.6e}$$

$$\gamma_{\mathrm{zx}} = \frac{\tau_{\mathrm{zx}}}{G} \tag{7.6f}$$

where E_{m} is Young's modulus (N/m^2), G is the shear modulus (N/m^2), and ν_{p} is Poisson's ratio (–).

Equations (7.6a–f) may also be written as:

$$\varepsilon_x = \frac{1}{3K}\left(\frac{\sigma_x + \sigma_y + \sigma_z}{3}\right) + \frac{1}{3G}\left[\sigma_x - \frac{1}{2}(\sigma_y + \sigma_z)\right] \tag{7.7a}$$

$$\varepsilon_y = \frac{1}{3K}\left(\frac{\sigma_x + \sigma_y + \sigma_z}{3}\right) + \frac{1}{3G}\left[\sigma_y - \frac{1}{2}(\sigma_x + \sigma_z)\right] \tag{7.7b}$$

$$\varepsilon_z = \frac{1}{3K}\left(\frac{\sigma_x + \sigma_y + \sigma_z}{3}\right) + \frac{1}{3G}\left[\sigma_z - \frac{1}{2}(\sigma_x + \sigma_y)\right] \tag{7.7c}$$

$$\gamma_{\mathrm{xy}} = \frac{\tau_{\mathrm{xy}}}{G} \tag{7.7d}$$

$$\gamma_{\mathrm{yz}} = \frac{\tau_{\mathrm{yz}}}{G} \tag{7.7e}$$

$$\gamma_{\mathrm{zx}} = \frac{\tau_{\mathrm{zx}}}{G} \tag{7.7f}$$

where K is the bulk modulus for the material (N/m^2).

E_{m}, G, K, and ν_{p} are related by:

$$E_{\mathrm{m}} = 2G(1 + \nu_{\mathrm{p}}), E_{\mathrm{m}} = 3K(1 - 2\nu_{\mathrm{p}})$$

The first terms on the right-hand side of Eqs. (7.7a)–(7.7c) are associated with change in volume of the element. This becomes evident when the three equations are added to give the volume change:

$$\varepsilon_x + \varepsilon_y + \varepsilon_z = \frac{1}{K}\left(\frac{\sigma_x + \sigma_y + \sigma_z}{3}\right)$$

The mean of the three normal stresses $\left(\frac{\sigma_x+\sigma_y+\sigma_z}{3}\right)$ is referred to as the hydrostatic stress. The second terms in the expressions are associated with change in shape.

To calculate stresses from measured elastic strains, the equations can be rewritten in the form:

$$\sigma_x = \frac{\nu_p E}{(1+\nu_p)(1-2\nu_p)}(\varepsilon_x+\varepsilon_y+\varepsilon_z) + \frac{E}{(1+\nu_p)}\varepsilon_x \tag{7.7g}$$

$$\sigma_y = \frac{\nu_p E}{(1+\nu_p)(1-2\nu_p)}(\varepsilon_x+\varepsilon_y+\varepsilon_z) + \frac{E}{(1+\nu_p)}\varepsilon_y \tag{7.7h}$$

$$\sigma_z = \frac{\nu_p E}{(1+\nu_p)(1-2\nu_p)}(\varepsilon_x+\varepsilon_y+\varepsilon_z) + \frac{E}{(1+\nu_p)}\varepsilon_z \tag{7.7i}$$

7.2.1.2 Plastic Flow of Metals In uniaxial loading such as occurs during a simple tensile test, plastic flow or permanent deformation is known to occur when the yield stress is exceeded. For a body that is subjected to a more general multiaxial stress state, a yield criterion is used to determine when plastic flow occurs. One that is commonly used is the von Mises yield criterion, which in simple terms states that

$$(\sigma_1-\sigma_2)^2+(\sigma_2-\sigma_3)^2+(\sigma_3-\sigma_1)^2 = C_k \tag{7.8}$$

where σ_1, σ_2, and σ_3 are the principal stresses acting on any element and C_k is a constant (Pa^2).

In a uniaxial tensile test, σ_1 is positive and $\sigma_2 = \sigma_3 = 0$. Furthermore, yielding occurs when $\sigma_1 = S_y$, where S_y is the yield strength of the material. Thus, from Eq. (7.8), $C_k = 2{S_y}^2$, giving

$$(\sigma_1-\sigma_2)^2+(\sigma_2-\sigma_3)^2+(\sigma_3-\sigma_1)^2 = 2{S_y}^2 \tag{7.9a}$$

or in more general terms,

$$(\sigma_x-\sigma_y)^2+(\sigma_y-\sigma_z)^2+(\sigma_z-\sigma_x)^2+6({\tau_{xy}}^2+{\tau_{yz}}^2+{\tau_{zx}}^2) = 2{S_y}^2 \tag{7.9b}$$

To facilitate analysis, an effective stress, $\overline{\sigma}$, is defined such that it has the same effect as the multiaxial stress state. Yielding or plastic flow occurs when its value reaches the yield strength, S_y. Thus, considering the von Mises yield criterion,

$$\overline{\sigma} = \frac{1}{\sqrt{2}}[(\sigma_1-\sigma_2)^2+(\sigma_2-\sigma_3)^2+(\sigma_3-\sigma_1)^2]^{1/2} \tag{7.10}$$

or

$$\overline{\sigma} = \frac{1}{\sqrt{2}}[(\sigma_x-\sigma_y)^2+(\sigma_y-\sigma_z)^2+(\sigma_z-\sigma_x)^2+6({\tau_{xy}}^2+{\tau_{yz}}^2+{\tau_{zx}}^2)]^{1/2} \tag{7.11}$$

Yielding thus occurs when

$$\overline{\sigma} = S_y \tag{7.12}$$

In linear elasticity, the current state of stress completely defines the state of strain. However, the current state of stress cannot define the plastic strains measured from the virgin (e.g. annealed) condition of the material. Any plastic deformation done before the current stresses were applied would have to be known before the total plastic strains could be defined. The value of $\overline{\sigma}$ necessary

to maintain plastic flow thus depends on the sum of the effective plastic strain increments accrued over the entire load path from the virgin condition of the material.

7.2.2 Plane Stress and Plane Strain

The discussion in Section 7.2.1 is based on a general three-dimensional analysis. To simplify subsequent discussions with respect to residual stresses, we outline the concepts of plane stress and plane strain, which are valid for a number of engineering applications.

7.2.2.1 Plane Stress An example of a body under plane stress conditions is a relatively thin plate which is subjected to a two-dimensional stress field. Consider a thin flat plate of uniform thickness which is loaded such that the applied loads are parallel to the plane of the plate and are uniformly distributed through its thickness (Figure 7.9). The state of stress on the top and bottom free surfaces are such that

$$\sigma_z = \tau_{zx} = \tau_{zy} = 0 \tag{7.13}$$

Since the plate is assumed to be thin, this condition is considered to hold throughout the plate. Thus, the only stresses that exist within the plate are σ_x, σ_y, and τ_{xy}, and are functions of x and y alone. These are the assumptions of generalized plane stress.

7.2.2.2 Plane Strain Let us now consider a prismatic cylinder which has the same cross-sectional area as the plate referred to in Section 7.2.2.1, and let it be subjected to the same uniform loading per unit length in the z-direction (Figure 7.10). However, its thickness in the z-direction is much greater. It is assumed that

1. Normal cross sections which were plane before loading remain plane after loading, that is, they do not distort in the z-direction. In other words, the displacement w in the z-direction is a linear function of z alone.
2. All normal cross sections well away from the ends undergo the same deformation in the x–y plane. In other words, the displacements u and v are functions of x and y alone.

These are the fundamental assumptions of plane strain and lead to the following equations:

$$\varepsilon_z = \text{constant at any } z, \gamma_{yz} = 0, \gamma_{zx} = 0 \tag{7.14a}$$

Thus, it can be shown that all the stresses are functions of x and y alone, and that

$$\tau_{zx} = \tau_{zy} = 0 \tag{7.14b}$$

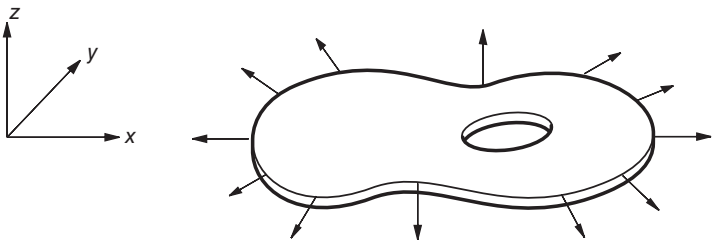


Figure 7.9 Plane stress state in a body.

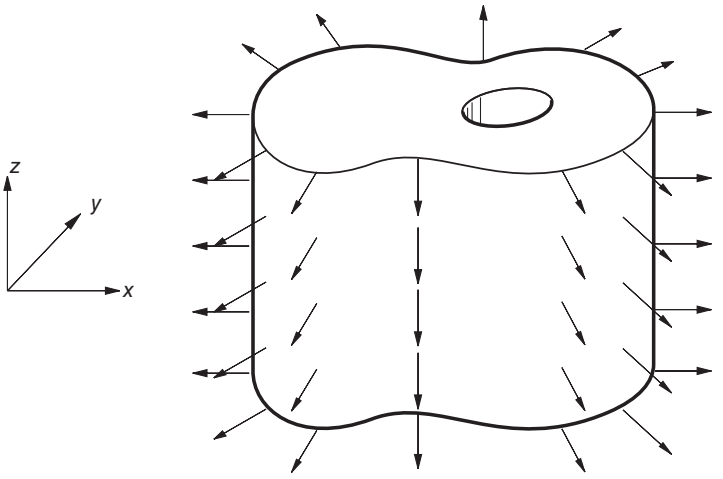


Figure 7.10 Plane strain state in a body.

7.2.2.3 Plane Stress/Plane Strain Equations Using the simplifications in the preceding sections, the equilibrium equations for both plane stress and plane strain are given by:

$$\frac{\partial \sigma_x}{\partial x} + \frac{\partial \tau_{xy}}{\partial y} = 0 \tag{7.15a}$$

$$\frac{\partial \tau_{xy}}{\partial x} + \frac{\partial \sigma_y}{\partial y} = 0 \tag{7.15b}$$

And the strain–displacement equations are

$$\varepsilon_x = \frac{\partial u}{\partial x} \tag{7.16a}$$

$$\varepsilon_y = \frac{\partial v}{\partial y} \tag{7.16b}$$

$$\gamma_{xy} = \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \tag{7.16c}$$

7.2.2.4 Compatibility Equation The three strain components in Eq. (7.16a–c), ε_x, ε_y, γ_{xy}, are functions of only two variables, u and v, and are therefore not independent. The compatibility equation which the total strains must satisfy is given by:

$$\frac{\partial^2 \varepsilon_x}{\partial y^2} + \frac{\partial^2 \varepsilon_y}{\partial x^2} = \frac{\partial^2 \gamma_{xy}}{\partial x \partial y} \tag{7.17}$$

Residual stresses exist when the nonelastic (plastic, thermal, etc.) strain components do not satisfy Eq. (7.17). Thus, an uneven distribution of plastic strains in a material results in residual stresses.

7.2.2.5 Stress–Strain Relations for Plane Stress/Plane Strain Here, we refocus our attention on elastic deformation. The stress–strain relations for generalized plane stress and plane strain are as follows:

Plane Stress:

$$2G\varepsilon_x = \sigma_x - \frac{\nu_p}{1+\nu_p}(\sigma_x + \sigma_y) \tag{7.18a}$$

$$2G\varepsilon_y = \sigma_y - \frac{\nu_p}{1+\nu_p}(\sigma_x + \sigma_y) \tag{7.18b}$$

$$G\gamma_{xy} = \tau_{xy} \tag{7.18c}$$

Plane Strain:

$$2G\varepsilon_x = \sigma_x - \nu_p(\sigma_x + \sigma_y) \tag{7.19a}$$

$$2G\varepsilon_y = \sigma_y - \nu_p(\sigma_x + \sigma_y) \tag{7.19b}$$

$$\sigma_z = E_m\varepsilon_z + \nu_p(\sigma_x + \sigma_y) \tag{7.19c}$$

$$G\gamma_{xy} = \tau_{xy} \tag{7.19d}$$

From Eqs. (7.18a–c) and (7.19a–d), it is obvious that a solution for stresses σ_x and σ_y under plane stress conditions can be applied to plane strain if ν_p is replaced by $\nu_p/(1-\nu_p)$.

Example 7.1 Residual stresses obtained by measuring elastic strains recovered after a part is cut are calculated for a plane stress, using Hooke's law as follows:

$$\sigma_x = -\frac{E_m}{1-\nu_p{}^2}(\varepsilon_x{}^e + \nu_p\varepsilon_y{}^e)$$

What is the reason for the negative sign in the expression?

Solution:

Because a region that is under tensile residual stress will contract when cut since the tensile stresses arise because that region is prevented from contracting.

7.3 EFFECTS OF RESIDUAL STRESSES

The two primary effects of residual stresses are as follows:

1. Apparent change in the strength of a part.
2. Distortion of the part.

7.3.1 Apparent Change in Strength

Consider a relatively large diameter steel rod that is heated into the austenite range for a long enough time for the temperature to become uniform throughout. If it is then quenched, tensile residual

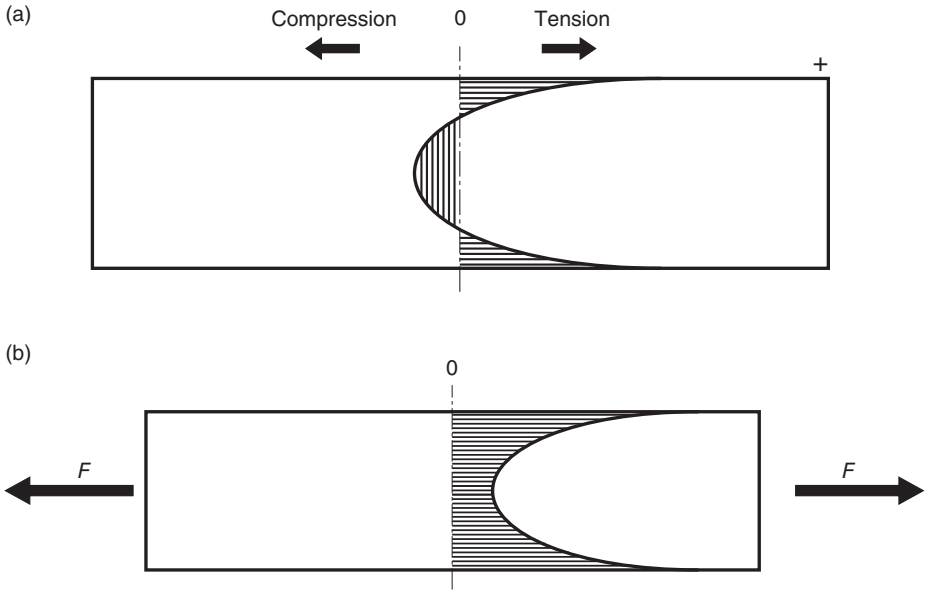


Figure 7.11 Residual stresses developed in a steel rod after quenching. (a) Possible residual stress distribution. (b) Stress state in the quenched bar after a tensile external load is applied.

stresses will be developed on the surface (Figure 7.11a), as a result of volume change associated with the austenite to martensite (density of austenite is higher than that of martensite) and other phase transformations that occur during quenching. Since tensile residual stresses then already exist in the rod after heat treatment, if it is subsequently loaded in tension, it will yield at applied stresses that are lower than the yield strength of the material. This is because the total stress on the rod surface will be the sum of the existing residual stresses and the applied stress (Figure 7.11b). In essence, in the elastic range, the residual stress can be added to the stress due to the applied external load. However, in the plastic region, the residual stress relaxes. This may result in a more uniform redistribution of the stresses (Figure 7.12).

If a glass plate is uniformly heated and the surfaces subsequently air-cooled, compressive residual stresses will be developed close to the surfaces (Figure 7.13), resulting in tempered glass. In this case, there is no phase transformation and thus no volume change during cooling. The mechanism is therefore similar to that discussed in Section 7.1.1. Such a plate can sustain higher bending moments, since the tensile stresses on the outer surface that result from bending will be reduced by the existing compressive residual stresses.

7.3.2 Distortion

When a material is subjected to heating and cooling during processing, thermal strains occur in the vicinity of the heat source. The stresses that result from these strains combine and react to produce internal forces that cause distortion, which is often evident as bending, buckling, and rotation (Figure 7.14). Principal changes that result in distortion of a plate, for example, include the following:

1. Transverse shrinkage perpendicular to the line along which the heat source moves (Figure 7.14a).
2. Longitudinal shrinkage parallel to the heat source line of motion (Figure 7.14b).

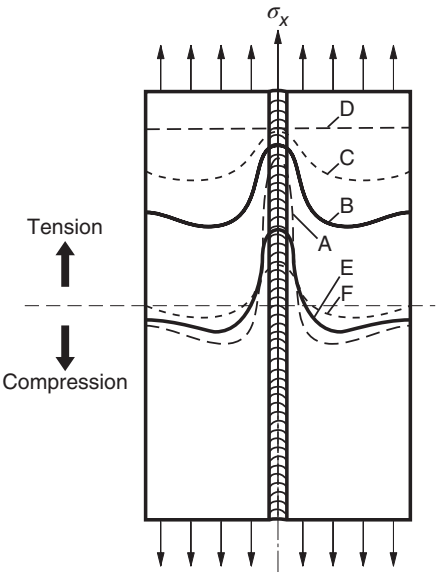


Figure 7.12 Redistribution of residual stresses in a weldment after loading into the plastic range. Source: Jenney and O'Brien (2001)/with permission from American Welding Society.

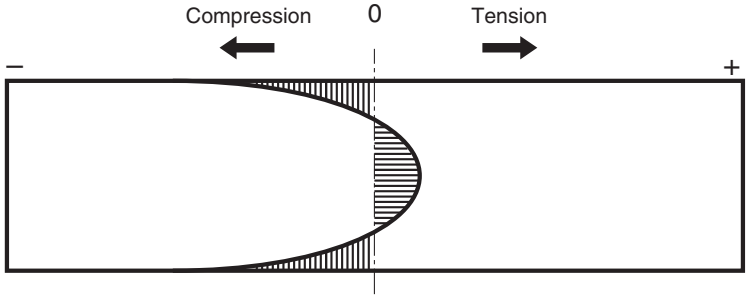


Figure 7.13 Residual stresses developed in a glass plate after uniform heating and subsequent air cooling of the surfaces.

3. Angular rotation about the heat source line of motion (Figure 7.14c). This is due primarily to nonuniformity of transverse shrinkage in the z-direction.
4. Buckling distortion (Figure 7.14d). Thin plates may be subjected to buckling instability when thermal compressive stresses are induced in them.

Actual distortion of structures tends to be more complex. Analysis of distortion is a coupled problem that involves three basic issues:

1. Analysis of the temperature field. This is essentially a three-dimensional conduction heat transfer problem involving a moving heat source.
2. Analysis of the plastic strains using the appropriate constitutive laws, incorporating temperature-dependent thermal expansion coefficient, elastic modulus, and flow stress. This would also involve a suitable yield criterion and restraint due to the geometry of the structure.

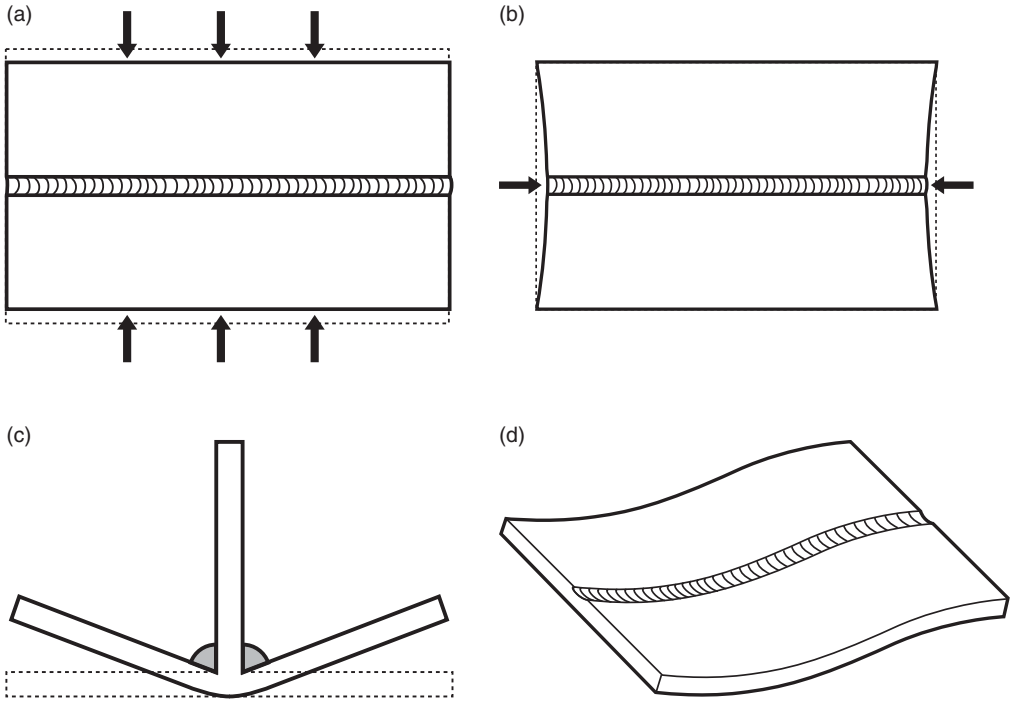


Figure 7.14 Various modes of distortion in weldments. (a) Transverse shrinkage. (b) Longitudinal shrinkage. (c) Angular change. (d) Buckling distortion. Source: Masubuchi (1980)/with permission of Elsevier.

3. Analysis of the distortion that results from the plastic strains. This involves consideration of force and moment equilibrium, strain hardening, and flow stress.

We now discuss some of the common methods used in measuring residual stresses.

7.4 MEASUREMENT OF RESIDUAL STRESSES

Various methods are available for measuring residual stresses. The major ones include

1. Stress relaxation.
2. X-ray diffraction.
3. Neutron diffraction.

Others are the ultrasonic and magnetic techniques. We shall limit our discussion to the three techniques listed. All three are based on the measurement of residual elastic strain. Residual stresses are then computed from the measured strains. The discussion on the measurement of residual stresses will be followed by a brief discussion on residual stress equilibrium, which can be used to check the validity of any measurements that are made.

7.4.1 Stress Relaxation Techniques

The stress relaxation techniques involve measurement of elastic strain recovered when the part is either cut into pieces or only parts are removed. Once the elastic strains are measured, the stresses

that were acting under those strains can be calculated using Hooke's law and considering the recovery or unloading process through the stress–strain curve (Figure 7.5). Using Eq. (7.18a–c) for the plane stress condition, we can express the residual stress in terms of the measured strains as:

$$\sigma_x = -\frac{E_{\mathrm{m}}}{1-{\nu_{\mathrm{p}}}^2}({\varepsilon_x}^{\mathrm{e}} + \nu_{\mathrm{p}}{\varepsilon_y}^{\mathrm{e}}) \tag{7.20a}$$

$$\sigma_y = -\frac{E_{\mathrm{m}}}{1-{\nu_{\mathrm{p}}}^2}({\varepsilon_y}^{\mathrm{e}} + \nu_{\mathrm{p}}{\varepsilon_x}^{\mathrm{e}}) \tag{7.20b}$$

$$\tau_{\mathrm{xy}} = -G\gamma_{\mathrm{xy}} \tag{7.20c}$$

The negative signs indicate that a part of the material that is subjected to tensile residual stresses will contract when relaxed, that is, residual stresses are removed. The strains are measured using strain gauges. The disadvantage with these techniques, however, is that they are destructive. However, they are very common and reliable methods. In this section, we discuss two primary types of the stress relaxation technique:

1. The sectioning technique.
2. The drilling technique.

The sectioning and drilling techniques are relatively rapid and less destructive, compared to other techniques such as the layer removal technique.

7.4.1.1 Sectioning Technique This is also sometimes referred to as the ring core method. With this method, electrical resistance strain gages are mounted on the surface of the structure. A circular disk piece that surrounds the strain gages is then removed by trepanning (Figure 7.15). The internal diameter of the piece removed may range between 15 and 150 mm, and the depth between 25% and 150% of its internal diameter. The change in strains after removal are measured. For a small enough disk, the removal may be assumed to have resulted in complete elastic recovery. Thus, the measured strains are the elastic strains. Residual stresses can then be calculated using Hooke's laws.

The sectioning technique is more sensitive than the drilling technique (Section 7.4.1.2), since it almost relieves the surface strains entirely. In addition, it is insensitive to small errors in the diameter or location of the annular hole relative to the strain gages. However, since the diameter is relatively large, the measured stresses are not very localized. Furthermore, more damage is done to the workpiece.

7.4.1.2 Drilling Technique The drilling technique is also known as the Mathar–Soete technique. In this case, strain gages are placed in a star or rectangular arrangement, and a small hole then drilled in the center of the arrangement (Figure 7.16). The hole diameter typically ranges between 1 and 4 mm and is drilled to a depth that is approximately equal to the diameter. Strain changes as measured by the three gages are then used to calculate the stresses.

This method is less sensitive, since the strains that are relieved decay quite rapidly with distance from the edge of the hole. Thus, the measured strains are only about 25–40% of the original residual strains at the hole location. Furthermore, the calculated stresses are significantly affected by errors in hole diameter or location. However, the hole drilling technique is more commonly used, since the measured stresses are more localized and less damage is done to the workpiece. In Section 7.4.1.3, we discuss how the measured strains can be converted to strains in cartesian coordinates.

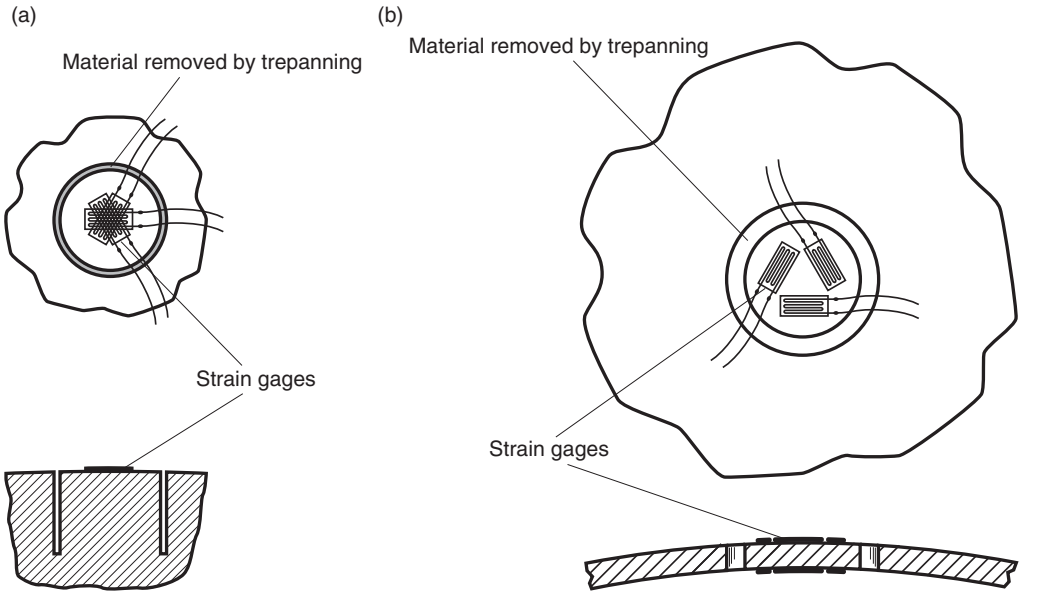


Figure 7.15 The sectioning method of residual stress measurement. (a) Thick plate. (b) Thin plate. Source: Reprinted from Lu (1996)/with permission from Fairmont Press, Inc.

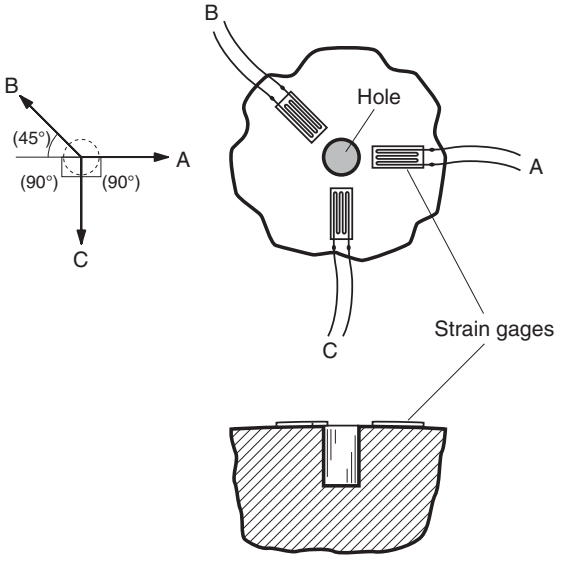


Figure 7.16 The drilling method of residual stress measurement. Source: From Lu (1996)/with permission from Fairmont Press, Inc.

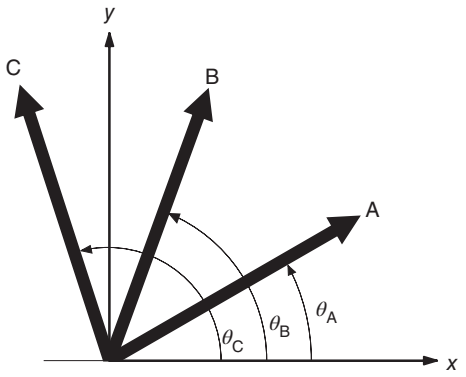


Figure 7.17 A general arrangement of strain gages.

7.4.1.3 Strain Analysis Consider the general case where strain gauges A, B, and C are mounted on the surface of a part at angles θ_A, θ_B, and θ_C with respect to the x-axis (Figure 7.17).

The strains ε_x, ε_y, and γ_{xy} are then related to those measured by the strain gages, ε_A, ε_B, and ε_C as:

$$\varepsilon_A = \varepsilon_x \cos^2\theta_A + \varepsilon_y \sin^2\theta_A + \gamma_{xy} \sin\theta_A \cos\theta_A \tag{7.21a}$$

$$\varepsilon_B = \varepsilon_x \cos^2\theta_B + \varepsilon_y \sin^2\theta_B + \gamma_{xy} \sin\theta_B \cos\theta_B \tag{7.21b}$$

$$\varepsilon_C = \varepsilon_x \cos^2\theta_C + \varepsilon_y \sin^2\theta_C + \gamma_{xy} \sin\theta_C \cos\theta_C \tag{7.21c}$$

ε_x, ε_y, and γ_{xy} can be calculated from Eq. (7.21a–c), from which the principal strains, ε_1 and ε_2 are obtained as:

$$\varepsilon_{1,2} = \frac{1}{2}(\varepsilon_x + \varepsilon_y) \pm \frac{1}{2}\sqrt{(\varepsilon_x - \varepsilon_y)^2 + {\gamma_{xy}}^2} \tag{7.22a}$$

$$\tan 2\varphi = \frac{\gamma_{xy}}{\varepsilon_x - \varepsilon_y} \tag{7.22b}$$

where ϕ is the angle between the maximum principal strain ε_1 (or maximum principal stress σ_1 if the material is isotropic) and the x-axis.

Knowing the measured strains, the residual stresses can be estimated using Eq. (7.20a–c).

Example 7.2 The residual stresses at a point in a thin titanium plate that has been formed using a laser are measured using a stress relaxation technique. The strain measurements obtained are shown in the figure below. Assuming Poisson's ratio to be 0.3, determine the maximum normal and shear stresses acting at that point.

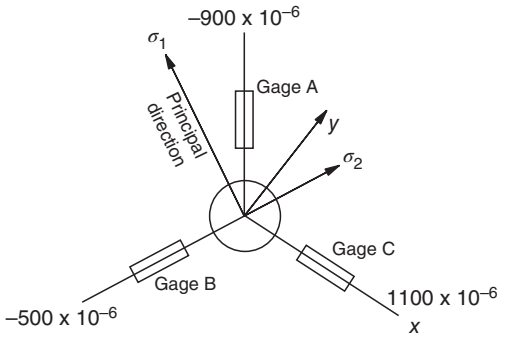


Figure E7.2

Solution:

From Appendix 7.B, Young's modulus for titanium is 116 GPa. To simplify things, we set up a coordinate system such that the x-axis is along Gage C. Then $\theta_C = 0°$, $\theta_A = 120°$, $\theta_B = 240°$. Thus,

$$\varepsilon_C = \varepsilon_x = 1100 \times 10^{-6}$$

Now using Eq. (7.21a–c), we have

$$\varepsilon_A = \frac{1}{4}(\varepsilon_x + 3\varepsilon_y - \sqrt{3}\gamma_{xy})$$
$$\varepsilon_B = \frac{1}{4}(\varepsilon_x + 3\varepsilon_y + \sqrt{3}\gamma_{xy})$$

Thus,

$$0.25 \times (1100 + 3\varepsilon_y - \sqrt{3}\gamma_{xy}) = -900$$
$$0.25 \times (1100 + 3\varepsilon_y + \sqrt{3}\gamma_{xy}) = -500$$

giving

$$\varepsilon_y = -1300 \times 10^{-6}, \qquad \gamma_{xy} = 462 \times 10^{-6}$$

Now from Eq. (7.22a), we obtain the principal strains as:

$$\varepsilon_{1,2} = \frac{1}{2}(1100 - 1300) \pm \frac{1}{2}\sqrt{(1100 + 1300)^2 + 462^2}$$
$$= \frac{1}{2}(-200 \pm 2444) = -1322 \quad \text{or} \quad 1122 \times 10^{-6}$$

The principal stresses are then obtained using Hooke's law:

$$\sigma_1 = -\frac{E_m}{1 - \nu_p{}^2}(\varepsilon_1 + \nu_p \varepsilon_2)$$
$$= -\frac{116 \times 10^9}{1 - 0.3^2}(1122 - 0.3 \times 1322) \times 10^{-6}$$
$$= -92.5 \text{ MPa} \quad \text{(minimum normal stress)}$$

$$\sigma_2 = -\frac{E_m}{1 - \nu_p{}^2}(\varepsilon_2 + \nu_p \varepsilon_1)$$
$$= -\frac{116 \times 10^9}{1 - 0.3^2}(-1322 + 0.3 \times 1122) \times 10^{-6}$$
$$= 125.6 \text{ MPa} \quad \text{(maximum normal stress)}$$

and the maximum shear stress is

$$\tau_{max} = \frac{\sigma_1 - \sigma_2}{2} = \frac{125.6 - (-92.5)}{2}$$
$$= 109.1 \text{ MPa} \quad \text{(maximum shear stress)}$$

7.4.2 X-ray Diffraction Technique

The X-ray diffraction technique is primarily used to measure surface stresses, up to a depth of about 50 μm. It is mainly for materials with a crystalline structure. The unit cell of each crystalline structure has lattice constants (lengths of the sides of the unit cell) which are characteristic of that material. When the material is subjected to elastic strains, that results in changes in the lattice constants. Plastic strains do not affect them since they result from dislocation motion. Thus, by measuring the lattice constants of the stressed material, the corresponding elastic strains can be determined. The lattice constants of the unstressed material are either known or are easily determined.

Measurement of the lattice constants is commonly made by X-ray diffraction. Advantages of this method are as follows:

1. It is capable of measuring strains in areas as small as 3 μm in diameter.
2. It is nondestructive.
3. Even though earlier measurement systems were very slow, more recent systems can make measurements in a few seconds or less.

However, it has the following disadvantages:

1. It is limited to surface stresses, up to about 2–50 μm from the surface.
2. Its accuracy is relatively low, especially for heat-treated parts with a distorted structure.

7.4.2.1 Principle of the X-ray Diffraction Technique This technique is based on Bragg's law. A beam of monochromatic radiation (AB and DF) incident on parallel atomic planes will be reflected in the direction BC and FH as shown in Figure 7.18. For the waves reflected from adjacent parallel atomic planes to reinforce each other, they must be out of phase by an integer (m) number of wavelengths. Thus, for diffraction, that is, reinforcement,

$$\text{EFG} = m\lambda \tag{7.23}$$

where λ is the wavelength of incident beam (X-ray) (μm).Or we can say that

$$\text{EF} = m\frac{\lambda}{2} \tag{7.24}$$

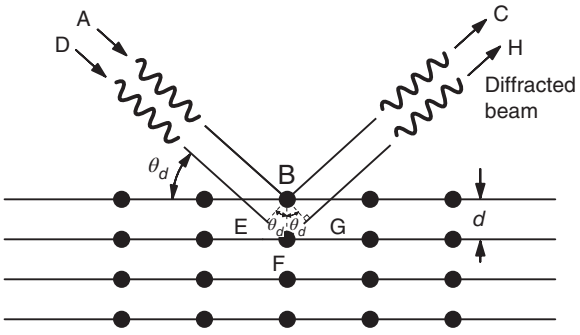


Figure 7.18 Principle of X-ray diffraction.

However,

$$\mathrm{EF} = d \sin \theta_d \tag{7.25}$$

where d is the inter-planar spacing and θ_d is the angle between incident beam and atomic plane.Therefore,

$$m\frac{\lambda}{2} = d \sin \theta_d \tag{7.26}$$

or

$$m\lambda = 2d \sin \theta_d \tag{7.27}$$

Equation (7.27) is known as Bragg's law. From it, the inter-planar spacing d can be calculated if λ and θ_d are known. Determining d also determines the lattice constants, a_{i}.

For a metal with a cubic lattice such as a face-centered cubic or body-centered cubic structure, the lattice constants are the same along all crystal axes, that is, $a_x = a_y = a_z = a$. Then, we have

$$d = \frac{a}{\sqrt{{h_{\mathrm{m}}}^2 + {k_{\mathrm{m}}}^2 + {l_{\mathrm{m}}}^2}} \tag{7.28}$$

where h_{m}, k_{m}, and l_{m} are the Miller indices of the reflecting plane.

h_{m}, k_{m}, and l_{m} incorporate the integer m and give a measure of the intersection of the reflecting plane on the crystal axes. In other words, m in Eq. (7.27) can be taken as 1. In reality, the Miller indices are the reciprocals of the intercepts of the plane with the x-, y-, and z-axes.

Constructive interference (diffraction) is obtained for an X-ray beam incident on the surface of a crystalline material when the beam meets lattice planes that are oriented in such a way that Bragg's law is fulfilled. Generally, a polycrystalline material will have some grains oriented suitably enough to produce a diffracted beam.

Let the lattice spacing for a fine-grained and stress-free polycrystalline specimen of a given material be d_0, and that for a stressed specimen of the same material be d. The elastic strain to which the specimen has been subjected can then be inferred from the change in crystal lattice as:

$$\varepsilon = \frac{d - d_0}{d_0} \tag{7.29}$$

Two of the methods on which the X-ray diffraction technique is based are as follows:

1. The film technique.
2. The diffractometer technique.

7.4.2.2 The Film Technique With this method, the X-ray beam is passed through a collimator and then directed toward the specimen (Figure 7.19). The reflected (or diffracted) beam is recorded using a film. When the film is developed, it shows near-circular rings, with each set of rings corresponding to an inter-planar distance, d_{hkl}. For each ring, θ_d from Bragg's law can be obtained using the following relation:

$$\frac{H_0}{l_{\mathrm{x}}} = 2 \tan(180 - 2\theta_d) \tag{7.30}$$

where l_{x} is the distance from the film to the specimen and H_0 is the diameter of the diffraction ring.

The strain is then obtained using Eq. (7.29). The film technique, though accurate, is relatively slow and thus not extensively used anymore.

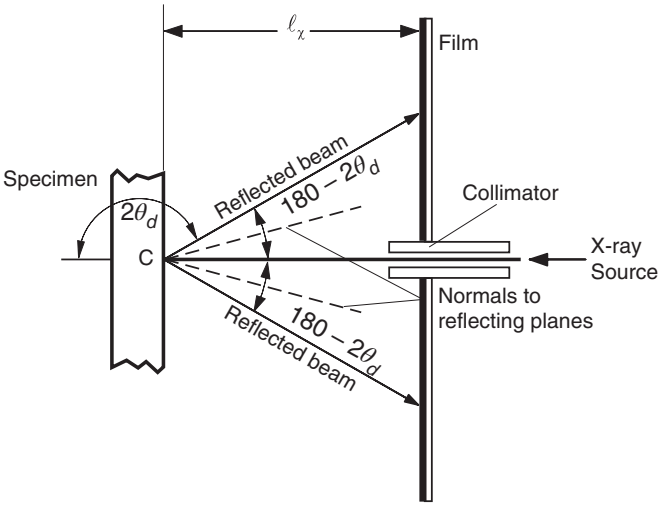


Figure 7.19 Schematic of the X-ray film technique. Source: Masubuchi (1980)/with permission of Elsevier.

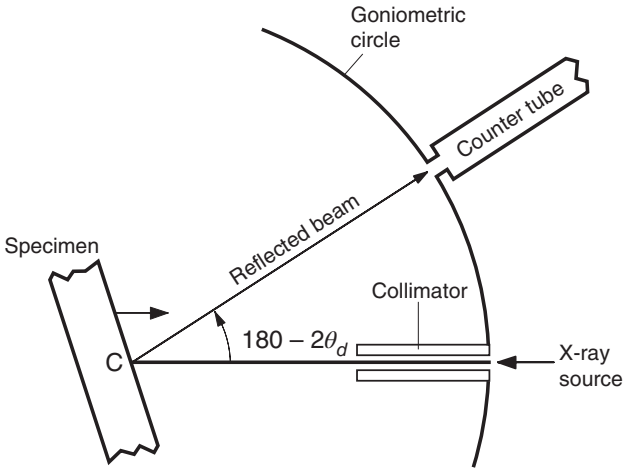


Figure 7.20 Schematic of the diffractometer technique. Source: Masubuchi (1980)/ with permission of Elsevier.

7.4.2.3 The Diffractometer Technique In this case, the diffraction angle is measured by moving a counter and receiving slit along a goniometer circle to measure the reflected beam intensity (Figure 7.20). The position of maximum beam intensity gives the diffraction angle. The major drawback of the diffractometer technique is that the specimen size is limited by equipment geometry. However, it is more accurate and quicker than the film technique. A measurement can be made in a few seconds or less. Thus, it is more widely used.

7.4.3 Neutron Diffraction Technique

The neutron diffraction technique is similar in principle to the X-ray diffraction technique. It is also based on the measurement of lattice strains. The main difference is that due to the ability

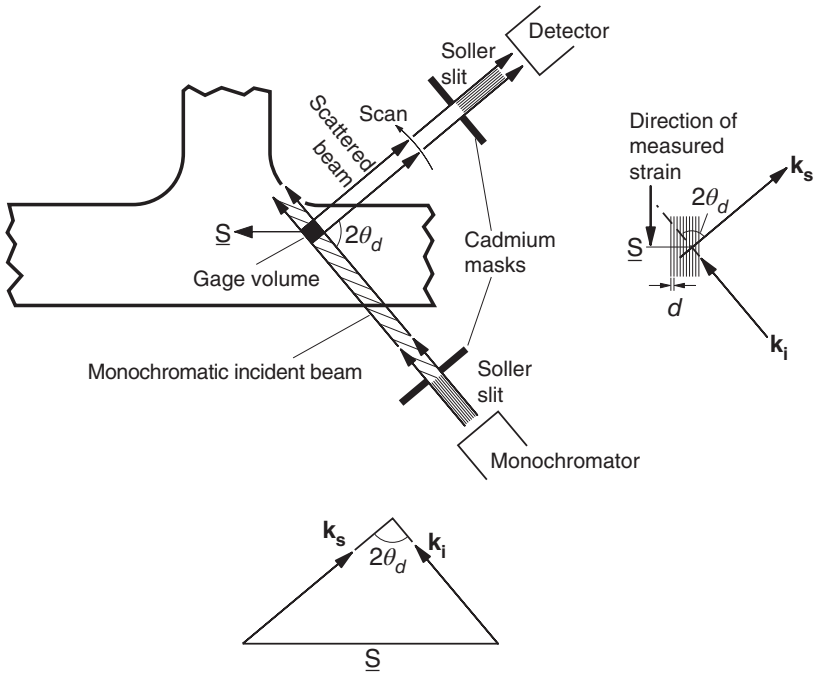


Figure 7.21 Principle of the neutron diffraction measurement technique. Source: Hutchings (1992)/with permission from Springer Nature.

of neutrons to penetrate more deeply into materials, neutron diffraction enables residual stresses to be measured much deeper into the material, that is, within the volume of the material. Even though the neutron beam may penetrate the entire volume, slits are used to define the local volume (Figure 7.21).

With this technique, a neutron beam from a reactor is first converted to a monochromatic beam of wavelength λ, using for example, a germanium crystal. A specific direction is imparted to the beam using a soller slit collimator. The cross-sectional area of the beam as it enters the sample and also as it enters the detector is determined by the aperture of neutron-absorbing cadmium masks (Figure 7.21), which are placed before and after the sample. The intersection of the incident and diffracted beams is referred to as the gage volume. Measurement is obtained of the average strain in grains that are properly oriented in a sample that is positioned completely within this volume. Larger samples may be moved through the gage volume to obtain the strain at different locations within it. The diffraction rings that result correspond to lattice planes that are oriented at an angle θ_d to the incident beam. These are also the planes whose normals lie along the scattering vector. The scattering vector, **s**, bisects the included angle between the incident and scattered beams. It is defined by the wave vectors of the incident and scattered beams:

$$\mathbf{s} = \mathbf{k_i} - \mathbf{k_s} \tag{7.31}$$

where $\mathbf{k_i}$ is the wave vector of the incident beam and $\mathbf{k_s}$ is the wave vector of the scattered beam.

When θ_d is obtained, the strain is then determined using Eq. (7.29). Measurements are usually made using either a crystal spectrometer or a time-of-flight diffractometer. A typical diffractometer is illustrated in Figure 7.22.

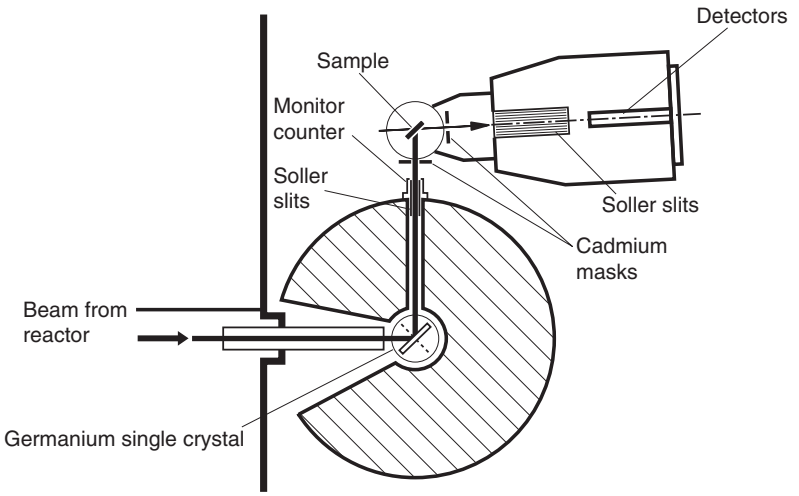


Figure 7.22 Schematic of a neutron diffractometer. Source: Hutchings (1992)/with permission from Springer Nature.

7.4.4 Residual Stress Equilibrium

Since residual stresses exist within a body even when there are no external loads acting on the body, the resultant force and moment of the residual stresses in a body must be zero. Thus, on any plane section in the body, the following equations must hold

$$\int \sigma \cdot dA = 0 \tag{7.32a}$$

$$\int dM = 0 \tag{7.32b}$$

where A is the area, M is the moment at any location, and σ is the stress at any point.

These equations can be used to validate experimentally determined residual stress data.

Example 7.3 The X-ray diffraction method is used to measure the residual stresses at a point in a thin steel plate subjected to laser heat treatment. The material has a lattice constant of 3.008×10^{-7} mm. The wavelength of the X-ray is 1.54×10^{-7} mm and is diffracted on the {1 0 1} planes. If uniaxial compressive stresses of 266 MPa are measured at a particular point, calculate the diffraction angle θ_d obtained during measurement in degrees.

Solution:

Young's modulus for steel is approximated as being the same as that for iron that, from Appendix 7.B, is given by $E_{\mathrm{m}} = 208$ GPa.

$$\sigma = E_{\mathrm{m}}\varepsilon$$
$$\Rightarrow -266 \times 10^6 = 208 \times 10^9 \times \varepsilon$$

or

$$\varepsilon = -1.279 \times 10^{-3}$$

Now

$$\varepsilon = \frac{a - a_0}{a_0} = \frac{a - 3.008}{3.008} = -1.279 \times 10^{-3}$$

Thus,

$$a = 3.00415 \times 10^{-7} \text{ mm}$$

But from Eq. (7.28),

$$d = \frac{a}{\sqrt{{h_\text{m}}^2 + {k_\text{m}}^2 + {l_\text{m}}^2}} = \frac{3.00415 \times 10^{-7}}{\sqrt{1^2 + 0^2 + 1^2}}$$
$$= 2.124 \times 10^{-7} \text{ mm}$$

Also, from Eq. (7.27),

$$m\lambda = 2d \sin \theta_d$$

Therefore,

$$1 \times 1.54 \times 10^{-7} = 2 \times 2.124 \times 10^{-7} \sin \theta_d$$

Giving

$$\theta_d = 21.26^\circ$$

7.5 RELIEF OF RESIDUAL STRESSES AND DISTORTION

Methods for relieving residual stresses and distortion can be grouped into two general categories:

1. Thermal treatments.
2. Mechanical treatments.

7.5.1 Thermal Treatments

Thermal treatments that are normally used to relieve residual stresses in welding, for example, include

1. Preheating.
2. Postheating.

7.5.1.1 Preheating Preheating involves heating the workpiece to a reasonably high temperature before it is processed. The preheating temperature depends on the workpiece material. For example, for ductile iron, this may be 650 °C, while for hardenable steels, it is normally 300 °C. A number of methods are available for preheating. These include

1. Heating in a furnace.
2. Induction coil heating.

3. Using a hot air blast.
4. Using another laser beam or splitting the original beam (Section 10.7.1.1).

The preheated material results in less contraction stresses since it is softer, and thus, plastic flow occurs more easily under the stresses that are developed, rather than shrinkage. Preheating also reduces cooling rates that result from the process, thereby minimizing the formation of hard phases.

However, it has the disadvantage that if the preheating temperature is too high, it may

1. Cause corrosion in stainless steels.
2. Soften the material being processed.
3. Reduce the strength of a workpiece that has already been hardened by heat treatment.

7.5.1.2 Postheating The use of postheating to relieve residual stresses is essentially a recovery/recrystallization process. The workpiece is heated to a temperature that depends on the material, for a specified time. Any residual stresses that exist in the material are reduced to the yield strength of the material at the treatment temperature. Subsequent cooling must be uniform throughout the workpiece to prevent the development of additional thermal stresses. If the postheating temperature is too high, undesirable properties may result. For carbon and low-alloy steels, stress relief is commonly performed between 1100 and 1350 °F. Postweld heat treatment is the most common method for relieving residual stresses in weldments.

7.5.1.3 Limitations of Thermal Stress Relief These include the following:

1. Thermal treatment of stainless steel weldments may result in sensitization and embrittlement as discussed in Section 5.1.3.3.
2. For low-carbon, high-strength steel weldments, reduction of strength and toughness in the HAZ, or hydrogen embrittlement, or both may occur when thermally treated.
3. Thermal treatment of bimetallic and multimetallic components can create problems.
4. Workpiece configurations with large variations in wall thickness may result in thermal stresses.
5. Some components may be too large for furnaces.
6. Asymmetrically shaped workpieces may result in nonuniform temperature distribution during cooling. Cooling rates can be slowed, but with increased costs.

7.5.2 Mechanical Treatments

Common mechanical treatments that are used in relieving residual stresses include

1. Peening.
2. Proof stressing.
3. Vibratory stress relief.

7.5.2.1 Peening Compressive residual stresses are induced in the surface layer of a material by applying impact loading. Peening has traditionally been done by directing a hammer or stream of small, high-velocity metal balls to the surface of the workpiece (shot peening). This is a cold working process. In recent years, lasers have been used to induce the same effect (laser shock peening,

[Section 11.5]). In addition to reducing residual stresses and distortion, peening may also improve the fatigue strength of the part, as a result of the compressive stresses that are induced. However, it has the disadvantage that it may conceal or introduce cracks and may thereby reduce the fracture toughness of the material.

7.5.2.2 Proof Stressing Proof stressing involves applying uniform loading to the entire workpiece. This results in a reduction of the residual stresses and a more uniform distribution after unloading (Figure 7.12). It serves as a means for stress relief and a test of structural integrity.

7.5.2.3 Vibratory Stress Relief This normally involves vibration of the part, usually at the natural frequency, for a period that ranges from 10 to 30 minutes depending on the weight of the workpiece. It is more economical than the thermal stress-relieving processes. The process induces yielding in parts of the structure and thereby relieves stresses and reduces distortion.

7.6 SUMMARY

Our discussion in this chapter started with the basic causes of residual stresses, where thermal stresses and nonuniform plastic deformation were identified as two of the major causes, with others being nuclear radiation and exposure to electromagnetic fields. The residual stresses that result during laser welding, for example, are thermal stresses and are tensile along the weld centerline, reducing to balancing compressive stresses as one moves further away from the centerline. This was followed by a basic discussion on stress analysis to lay the analytical framework for understanding residual stresses. This included discussions on stress–strain relations and conditions of plane stress and plane strain. Both elastic and plastic behaviors were presented, culminating in the compatibility equation, which determines the conditions under which residual stresses would exist in a body. In linear elasticity, the current state of stress completely defines the state of strain. However, the current state of stress cannot define the plastic strains measured from the virgin (for example, annealed) condition of the material. Any plastic deformation done before the current stresses were applied would have to be known before the total plastic strains could be defined.

The impact that residual stresses have on a part were then discussed, focusing on two primary effects: apparent change in strength and distortion. Depending on the nature of the residual stresses (compressive or tensile), and the external loading that is applied, the strength of the part may either be enhanced or reduced as the stresses due to external loading are superimposed on the existing residual stresses. Following this, some of the major methods for measuring residual stresses were discussed, focusing on the stress relaxation, X-ray diffraction, and neutron diffraction methods. Under stress relaxation, which is destructive, the sectioning and drilling techniques were discussed. Under X-ray diffraction, the basic principles leading up to the film and diffractometer techniques were discussed, where the latter was identified as being the most commonly used method due to its accuracy and speed of implementation. Since residual stresses exist within a body even when there are no external loads acting on it, the resultant force and moment of the residual stresses in the body must be zero. This may be used to check the efficacy of measurements that are made. Finally, different methods for relieving residual stresses in a part were outlined, with the primary methods being identified as either thermal or mechanical based. The thermal methods include preheating and postheating, while the mechanical methods include peening, proof stressing, and vibratory stress relief.

APPENDIX 7.A

List of symbols used in the chapter

Symbol	Parameter	Units
a, a_i	Lattice constants	m
C_k	Constant	Pa^2
d	Inter-planar spacing of a crystal structure	m
d_0	Inter-planar spacing of undeformed material	m
h_m, k_m, l_m	Miller indices of a reflecting plane	-
H_0	Diameter of diffraction ring	m
$\mathbf{k_i}$	Wavevector of incident beam	/m
$\mathbf{k_s}$	Wavevector of scattered beam	/m
K	Bulk modulus	Pa (N/m^2)
l_x	Distance from film to specimen	m
L_0	Original length of specimen	m
ΔL	Extension or contraction	m
M_{init}	Moment required to achieve yielding in outer fibers	N m
M_y	Moment required to achieve full yield	N m
$\mathbf{s}$	Scattering vector	/m
ΔT	Temperature change	°C
ν_p	Poisson's ratio	—
ϕ	Angle between maximum principal stress and x-axis	°
$\overline{\sigma}$	Effective stress	N/m^2
θ_d	Diffraction peak corresponding to lattice spacing d	°

APPENDIX 7.B

Mechanical properties of some pure annealed materials at 20 °C

Material	Young's modulus (GPa)	Yield strength (MPa)	Tensile strength (MPa)	Percent elongation (%)
Aluminum (Al)	62	15–20	40–50	50–70
Chromium (Cr)	248	—	83	0
Copper (Cu)	128	33.3	209	60
Gold (Au)	78	—	103	30
Iron (Fe)	208	276	414	18
Lead (Pb)	14	—	—	—
Magnesium (Mg)	44	21	90	2–6
Manganese (Mn)	191	241	496	40
Molybdenum (Mo)	324	≈400	≈500	—
Nickel (Ni)	207	59	317	30

Material	Young's modulus (GPa)	Yield strength (MPa)	Tensile strength (MPa)	Percent elongation (%)
Niobium (Nb)	103	207	275	30
Silicon (Si)	106–113	—	—	—
Silver (Ag)	76	—	125	—
Tin (Sn)	43	—	—	—
Titanium (Ti)	116	140	235	54
Tungsten (W)	345	—	—	—
Vanadium (V)	131	—	—	—
Zinc (Zn)	92	—	—	—
Zirconium (Zr)	94	≈230	—	20–35

Source: Adapted from Bardes (1978a,b), Barrett et al. (1973), Felbeck and Atkins (1996), MatWeb.

PROBLEMS

7.1. Given the strain–displacement equations

$$\varepsilon_x = \frac{\partial u}{\partial x}$$

$$\varepsilon_y = \frac{\partial v}{\partial y}$$

$$\gamma_{xy} = \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}$$

Derive the compatibility equation

$$\frac{\partial^2 \varepsilon_x}{\partial y^2} + \frac{\partial^2 \varepsilon_y}{\partial x^2} = \frac{\partial^2 \gamma_{xy}}{\partial x \partial y}$$

7.2. Show that a solution for stresses σ_x and σ_y under plane stress conditions can be applied to plane strain if ν_p is replaced by $\nu_p/(1 - \nu_p)$.

7.3. Starting with the strain compatibility equation, obtain the stress compatibility equation:

$$\frac{\partial^2}{\partial y^2}(\sigma_x - \nu_p \sigma_y) + \frac{\partial^2}{\partial x^2}(\sigma_y - \nu_p \sigma_x) - 2(1 + \nu_p)\frac{\partial^2 \tau_{xy}}{\partial x \partial y} = 0$$

7.4. Obtain expressions for ε_x, ε_y, and γ_{xy} in terms of ε_A, ε_B, and ε_C for the delta rosette shown as follows.

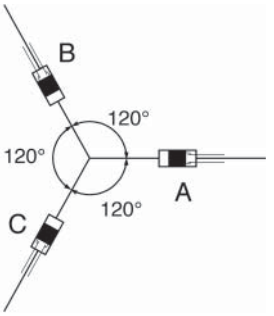


Figure P7.4

7.5. Obtain expressions for ε_x, ε_y, and γ_{xy} in terms of ε_A, ε_B, and ε_C for the rectangular rosette shown as follows.

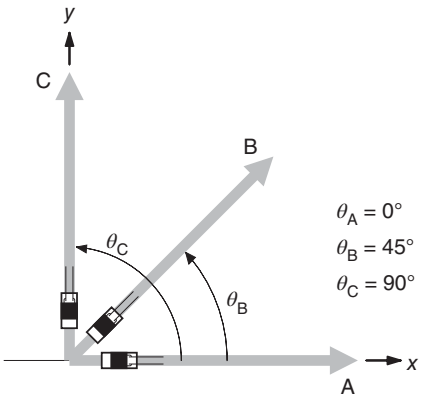


Figure P7.5

Obtain expressions for the principal stresses, σ_1 and σ_2, in terms of ε_A, ε_B, and ε_C.

7.6. The middle member (2) of the fixture shown as follows is heated to a temperature close to its melting temperature, about 100 °C below the melting temperature. Plot the variation of stress

versus temperature in member 1, while member 2 is first being heated and then cooled back to room temperature. Label the heating and cooling cycles.

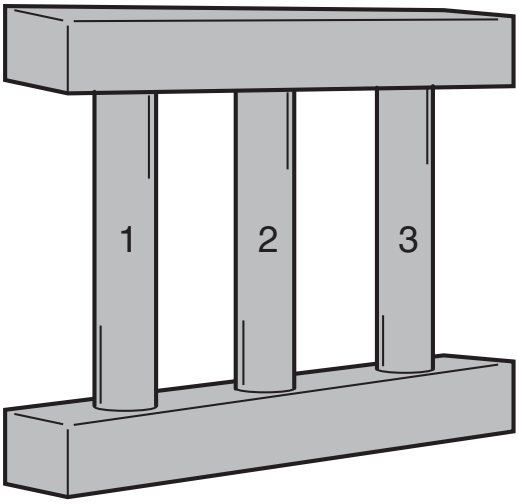


Figure P7.6

7.7. (a) A rectangular strain gauge rosette arranged as shown in the figure below is used to measure the residual stresses in a weldment. If the strain gauges measure strains in the directions **A, B, and C,** express the strains ε_x, ε_y, and γ_{xy} in terms of the measured strains ε_A, ε_B, and ε_C.

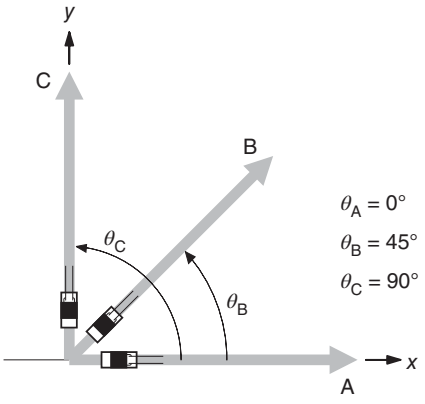


Figure P7.7

(b) If ε_A, ε_B, and ε_C are respectively, 1000×10^{-6}, -200×10^{-6}, and 750×10^{-6}, determine the maximum normal stress at that point if Young's modulus for the material is 200 GPa and the Poisson ratio is 0.3.

7.8. (a) After a part is processed using a laser, a small square piece is cut from a region which is subjected to compressive residual stresses. Will the piece expand or contract after it is cut? Why?

(b) Explain why preheating would reduce distortion in a laser-welded part.

7.9. The residual stresses at a point in an aluminum plate that has been subjected to some form of laser treatment are measured by the X-ray diffraction method. The X-rays, of wavelength 1.54×10^{-7} mm, are diffracted $2\theta_d = 38.4°$ on the {1 1 1} planes. If the lattice constant of undeformed aluminum is 4.049×10^{-7} mm, estimate the residual stresses that exist at that point. Consider the stresses to act in only one direction.

7.10. A consultant claims to have developed a new method for measuring residual stresses using a three-element rectangular rosette as shown in the figure. If the strains measured by the gages A, B, and C are −500, 300, and 150 μm/m respectively, calculate the maximum stress at the point being measured, if the material is steel with a Poisson ratio of 0.3.

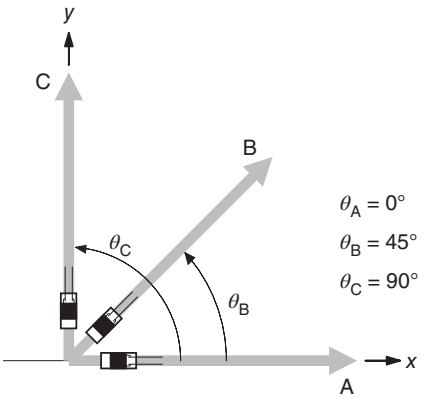


Figure P7.10

7.11. In measuring the residual stresses in a thin copper plate that has been subjected to laser forming using the X-ray diffraction method, the diffraction angle is obtained as 42° for the plane with the Miller indices {1 1 1}. If the lattice constant of copper is 3.615×10^{-7} mm, what is the wavelength of the X-rays used?

7.12. The drilling technique is used to measure the residual stresses at a point in a laser-welded thin steel plate. The strain measurements obtained are shown in the figure below.

Assuming a Poisson ratio of 0.3, determine the following stresses that act at that point:

1. The maximum normal stress.
2. The maximum shear stress.

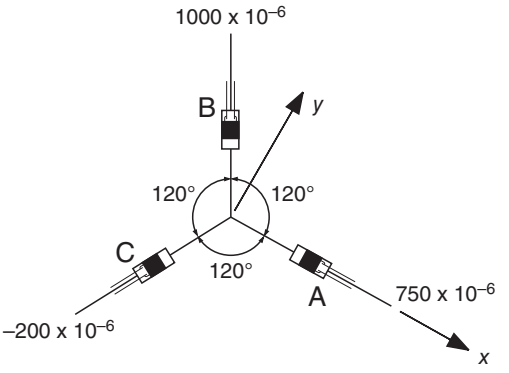


Figure P7.12

7.13. A beam made from an elastic-perfectly plastic material is subjected to a bending moment until the entire cross section is plastically deformed. Show that on elastic recovery, the elastic spring-back stress distribution will be as shown by the dashed line in the figure below.

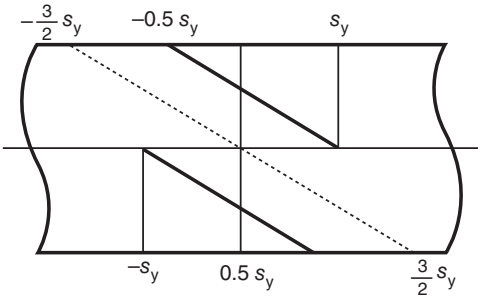


Figure P7.13

7.14. Assume that the distribution of longitudinal residual stress σ_x in a weldment can be approximated by the following equation:

$$\sigma_x = \mathrm{A}\left[1 - \left(\frac{y}{\mathrm{B}}\right)^2\right] e^{-\frac{1}{2}\left(\frac{y}{\mathrm{B}}\right)^2}$$

(a) Explain the physical meaning of the coefficients A and B in the equation.
(b) Label A and B in a sketch of the longitudinal stress distribution.
(c) Based on this equation, what is the residual stress σ_x, when $y = 0$, B, 1.5B, 5B?

7.15. The residual stress in a thin copper plate is measured using the X-ray diffraction method. The X-ray, of wavelength 4.492×10^{-7} mm, is diffracted into several circular rings. One of the ring diameters is 23.77 mm when the distance from the specimen to the film is 15 mm. Assuming the lattice constant of undeformed copper to be 3.6149×10^{-7}mm,

Calculate the residual stress at the point of interest, assuming $m = 1$.

(a) Discuss if $m = 1$ is a proper assumption.
(b) Explain why there would be other rings if there is only one θ_d corresponding to each ring diameter. Suggest another possible ring diameter.

BIBLIOGRAPHY

Bardes, B.P. (ed.) (1978a). *Metals Handbook*, *Properties and Selection*: *Irons and Steels*, 9e, vol. 1. OH: American Society for Metals.

Bardes, B.P. (ed.) (1978b). *Metals Handbook*, *Properties and Selection*: *Nonferrous Alloys and Pure Metals*, 9e, vol. 2. OH: American Society for Metals.

Barrett, C.R., Nix, W.D., and Tetelman, A.S. (1973). *The Principles of Engineering Materials*. Englewood Cliffs, NJ: Prentice Hall.

Dally, J.W. and Riley, W.F. (1991). Strain-analysis methods. In: *Experimental Stress Analysis*, 3e, 311–340. NY: McGraw Hill, Inc.

Felbeck, D.K. and Atkins, A.G. (1996). *Strength and Fracture of Engineering Solids*, 2e. Englewood Cliffs, NJ: Prentice Hall.

Holden, T.M. and Roy, G. (1996). The application of neutron diffraction to the measurement of residual stress and strain. In: *Handbook of Measurement of Residual Stresses* (ed. J. Lu), 133–178. Lilburn, GA: The Fairmont Press, Inc.

Hutchings, M.T. (1992). Neutron diffraction measurement of residual stress fields: overview and points for discussion. In: *Measurement of Residual and Applied Stress Using Neutron Diffraction* (ed. M.T. Hutchings and A.D. Krawitz), 3–18. Dordrecht, Netherlands: Kluwer Academic Publishers.

Jenney, C.L. and O'Brien, A. (ed.) (2001). *Welding Handbook*, 9e. Miami, FL: American Welding Society.

Lu, J. (ed.) (1996). *Handbook of Measurement of Residual Stresses*. Lilburn, Georgia: The Fairmont Press, Inc.

Ludema, K. (1998). *Unpublished Manual*. Ann Arbor, MI.

Masubuchi, K. (1980). *Analysis of Welded Structures*. Oxford, England: Pergamon Press.

Noyan, I.C. and Cohen, J.B. (1987). *Residual Stress – Measurement by Diffraction and Interpretation*. New York: Springer-Verlag.

Schajer, G.S., Roy, G., Flaman, M.T., and Lu, J. (1996). *Hole-drilling and ring core methods, Handbook of Measurement of Residual Stresses* (ed. J. Lu), 5–34. Lilburn, Georgia: The Fairmont Press, Inc.

Soete, W. (1949). Measurement and relaxation of residual stresses. *Welding Journal* 28: 354s–364s.

Webster, G.A. (1992). Role of neutron diffraction in engineering stress analysis. In: *Measurement of Residual and Applied Stress Using Neutron Diffraction* (ed. M.T. Hutchings and A.D. Krawitz), 21–35. Dordrecht, Netherlands: Kluwer Academic Publishers.

PART III
Laser Materials Processing

8 Background on Laser Processing

The laser beam is the heat source in laser materials processing. Even though the laser is normally considered to be a light source, it is also a form of energy, and as such, can be a useful source of intense heat when concentrated by focusing. Lasers are able to produce high-energy concentrations because of their monochromatic, coherent, and low divergence properties as compared to an ordinary light source. As a result, they can be used to heat, melt, and vaporize most materials. The processes for which lasers are commonly used include welding, cutting, surface modification (including heat treatment), forming, and additive manufacturing. The power densities and exposure times necessary for various processes are shown in Figure 8.1. To provide a frame of reference, a comparison of various heat sources is made in Table 8.1. For the laser beam, the highest power densities are achieved with the beam in the Gaussian (TEM_{00}) mode (Section 8.1.3).

In this chapter, we start with a general discussion on system-related parameters that have the most significant effect on process output. The major stages of energy loss from the power drawn from the mains supply to the actual power used in the process are then discussed in terms of process efficiency. This is followed by a discussion on disturbances that affect process quality, and finally, the advantages and disadvantages of laser processing in general are briefly presented.

8.1 SYSTEM-RELATED PARAMETERS

The principal system parameters that affect process quality include the beam power and power density at the focal point, wavelength and focusing system, beam mode (distribution), beam form (continuous wave (CW) or pulsed), and beam quality. Others include beam absorption, beam delivery, beam alignment, and the motion unit used for positioning the beam or workpiece. Our discussion follows in that order. Some parameters such as beam polarization, assist and/or shielding gas, focal point location, and traverse speed are more process-specific and are thus discussed in relation to individual processes in subsequent chapters.

8.1.1 Power and Power Density

The power density at the focal point is determined by the incident power and the size of the focused beam. For a Gaussian distributed laser beam, the beam size is defined as the radial distance at which the intensity of the beam reduces to $1/e^2$ or 13.5% of the peak intensity. This is also the area which has approximately 87.5% of the total beam energy. The power density, PD, is given by:

$$\mathrm{PD} = \frac{q}{\pi w^2} \tag{8.1}$$

where q is the power of the laser beam (W) and w is the beam radius (mm).

Principles of Laser Materials Processing: Developments and Applications, Second Edition. Elijah Kannatey-Asibu, Jr.
© 2023 John Wiley & Sons, Inc. Published 2023 by John Wiley & Sons, Inc.
Companion Website: www.wiley.com/go/Kannatey-Asibu/PrinicplesLaserMaterialsProcessing

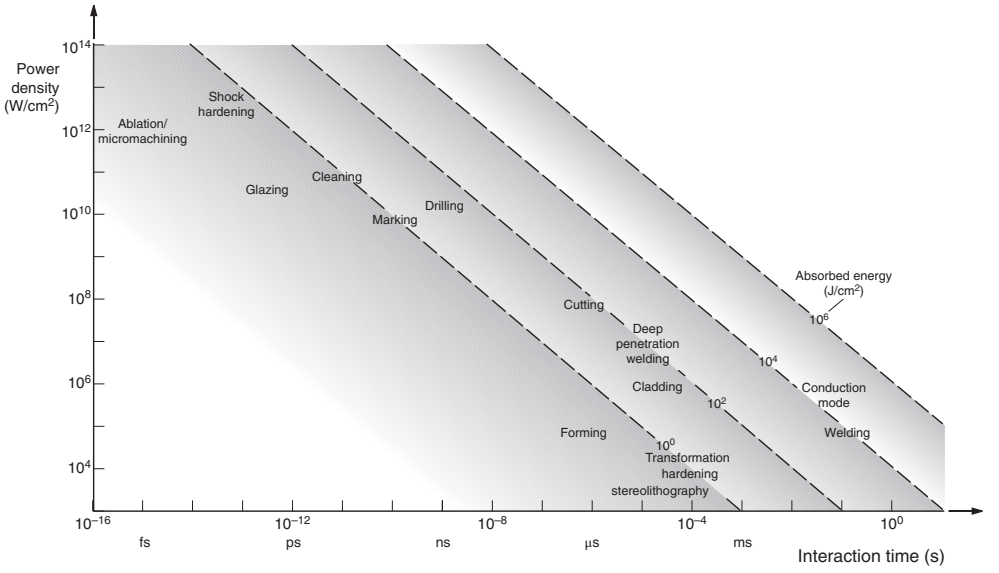


Figure 8.1 Power densities and interaction times for various laser processes.

Table 8.1 Comparison of power densities for various heat sources.

Heat source	Power density (W/cm^2)
Blackbody radiation (3000 °C)	6.45×10^4
Oxyacetylene flame	10^5
Transferred arc plasma torch	10^7
Electron beam	10^{11}
Laser (continuous CO_2)	2.5×10^8
Laser (pulsed Nd:YAG)	$>10^{14}$
Laser (ultrashort pulse)	$>10^{21}$

8.1.2 Wavelength and Focusing

The wavelength of the beam influences the process in several ways, including absorption of the beam and the extent of focusing. In this section, we shall discuss its influence on focusing. Its effect on beam absorption will be separately addressed in Section 8.1.6.

When a Gaussian beam of radius w_0 and wavelength λ is incident on a lens of focal length f_l, it is focused to a radius w_f given by:

$$w_f = \frac{1}{\pi}\left(\frac{f_l \lambda}{w_0}\right) \tag{8.2}$$

where w_0 is the incident beam radius at the lens.

Equation (8.2) indicates that the power density at the focal point is inversely proportional to the square of the wavelength. Thus, the smaller the wavelength of the beam, the higher the output power density, all other conditions being equal. The shorter wavelength of the Nd:YAG laser (1.06 μm) thus gives it some advantages over the CO_2 laser (10.6 μm). For the same input power, beam mode, beam size, and optics, the power density ratio of an Nd:YAG laser to a CO_2 laser is 100. Equation (8.2) also

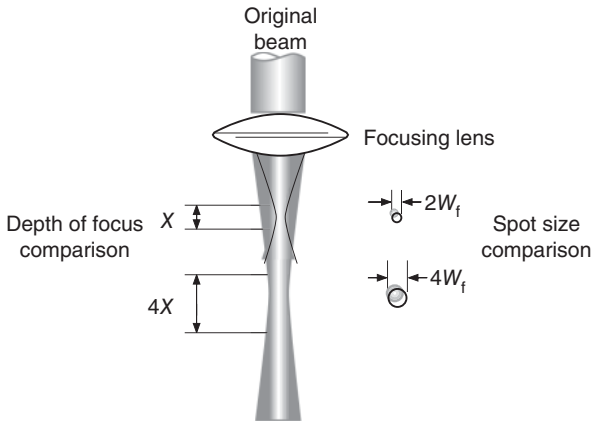


Figure 8.2 Lens focal length and corresponding depth of focus.

indicates that the greater the original beam diameter, the smaller the focused beam radius, and thus the greater the power density of the focused beam. Therefore, an expander is often used to increase the diameter of the beam incident on the focusing lens.

It is also evident from Eq. (8.2) that the shorter the focal length of the lens, the smaller is the size of the focused beam, and thus the greater the power density. Thus in laser welding, the use of a shorter focal length lens results in deeper penetration for the same beam power, especially at lower power levels. At much higher power levels of say, 20 kW, there is little difference, if any, on the impact of the focal length on penetration. The use of very short focal length lenses, however, has two disadvantages:

1. The depth of focus is correspondingly small (Figure 8.2), requiring greater precision in maintaining the lens to workpiece distance.
2. In welding, the shorter focal length lens is more liable to be damaged by vapor and molten metal ejected from the weld.

Focal lengths normally used range from 2.5 in. (65 mm) to 10 in. (254 mm), but 7.5 in. (190 mm) lenses are about the standard. Common lens materials include ZnSe, KCl, GaAs, and CdTe, and a typical beam radius at the focal point would be 100 μm. For very long focal lengths (say, 200 mm and over), a reflective parabolic copper mirror is often used for focusing.

The depth of focus, z_f, of a focusing system is the distance over which the workpiece can be moved from the original focal point without a significant change in the focused beam radius (Figure 8.2). A long depth of focus is good for laser cutting, for example, where parallel-sided cuts or holes are needed.

The depth of focus, z_f, over which 80% of the power density at the focal point is maintained can be shown to be given approximately (Problem 8.2) by:

$$z_f = \frac{\pi \times {d_f}^2}{4 \times \lambda} \tag{8.3}$$

where d_f is the focused beam diameter.

8.1.3 Beam Mode

In simple terms, the beam mode or profile is the intensity (irradiance or power per unit area) distribution in a plane perpendicular to the beam propagation axis. It is an indication of how the energy intensity is distributed over the beam cross section. The profile varies with propagation distance from the output mirror. The near-field profile may be distorted due to diffraction effects, but the profile tends to assume a constant, symmetrical shape in the far-field. Far-field is reliably detected after five Rayleigh ranges. The Rayleigh range (also known as the near-field distance), x_R, is defined as the distance between the beam waist and the point where the beam size is $\sqrt{2}$ times the waist size:

$$x_R \approx \frac{\pi {w_w}^2}{\lambda}$$

Distances within the Rayleigh range are referred to as the near-field region, while those much greater than the Rayleigh range are referred to as the far-field region. Depending on the type of beam, the Rayleigh range can be as little as a few millimeters, or as much as a number of meters.

Profiles of a Gaussian (TEM_{00}) (low order) mode and a multimode (TEM_{mn}) (high order) are shown schematically in Figure 8.3. The Gaussian or low-order mode (Figure 8.3a) can be focused to the laser's theoretical minimum radius. The higher order or multimode beams, on the other hand, are more spread out and are not as well focused, resulting in larger focal spot sizes and thus lower power density for the same output power. These are more common with higher power lasers.

A Gaussian mode beam is thus normally preferred for cutting, for example, since it results in the smallest possible focused beam radius and thus highest power density for a given incident power. On the other hand, for butt welding, a higher-order multimode beam is preferable since the larger beam diameter means less stringent joint fit-up requirements.

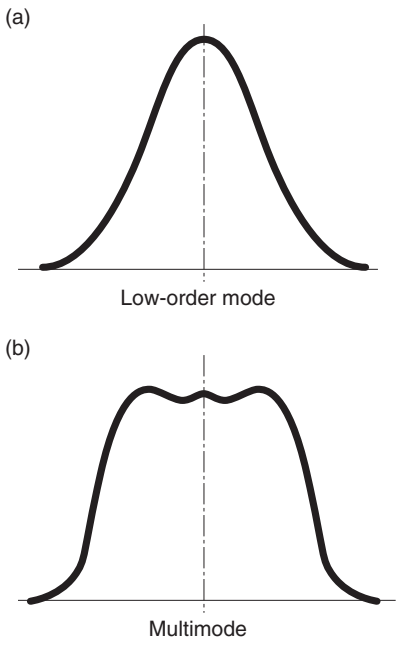


Figure 8.3 Beam distributions. (a) Low-order (Gaussian) mode. (b) High-order (multimode) mode.

Nd:YAG lasers normally generate a multimode beam (Figure 8.3b), since that ensures the most efficient power extraction from the rod used in the laser source. CO_2 lasers, on the other hand, can generate either a low-order or multimode beam, with the low-order mode being more common at power levels less than 5 kW. However, the output of high-power CO_2 lasers is normally multimode, since the high-power lasers are usually transverse flow machines which inherently produce multimode beams.

8.1.4 Beam Form

The laser output is usually of two forms:

1. *CW*: where the output power is essentially constant.
2. *Pulsed*: where the output power varies periodically, in a manner similar to a clock signal.

The CW output is used quite often. However, the pulsed output of a laser has advantages in a number of situations. For example, variation of the on/off ratio or duty cycle of the pulsed beam enables applications such as drilling and cutting of sharp-edged contours to be effectively controlled. The reduced total heat input also makes the pulsed output suitable for parts that are sensitive to distortion. On the other hand, the lower overall average power of the pulsed beam results in a reduced processing speed. Furthermore, in processes such as cutting and seam welding, where there is a need for continuity, the pulse repetition rate, when a pulsed beam is used, has to be high enough for the pulses to overlap.

Pulsed beams can be classified into three forms (Figure 8.4):

- *Gated pulsing*: which is obtained by switching the discharge current on and off to produce a pulsed output (Figure 8.4a). The peak output power is about the same as the nominal CW output power. The average power is the product of pulse energy and repetition rate, while the pulse energy depends on the peak power, pulse shape, and pulse duration.
- *Superpulsing*: where the peak output power is about two to three times that of the CW output (Figure 8.4b). However, the duty cycle is correspondingly reduced. Superpulsing enables low- to medium-power CW lasers to produce relatively high output powers over short periods of time and is similar to Q-switching except that it is achieved electronically.
- *Hyperpulsing*: where high peak power pulses are superposed on the CW output (Figure 8.4c).

Both superpulsing and hyperpulsing are useful for minimizing dross formation at the lower surface during cutting of some materials such as stainless steels. This is because much higher temperatures are achieved in the molten metal when either superpulsing or hyperpulsing is used. That results in a lower viscosity of the material, making it easier to eject with the same level of gas pressure. Both pulsing modes enable deeper weld penetration or greater cut thickness (up to about 50%) to be achieved for the same average power as a CW output. They also make it easier to process highly reflective and/or high thermal conductivity materials.

8.1.5 Beam Quality

The quality of a laser beam is normally characterized in terms of the K and M^2 factors, which are dimensionless, beam parameter product (BPP) with units mm-mrad, and beam divergence, θ. They indicate the resemblance of the beam to a Gaussian beam. These are related as follows:

$$M^2 = \frac{1}{K} = \text{BPP} \times \frac{\pi}{\lambda} = \frac{\pi w_w \theta}{\lambda} \tag{8.4}$$

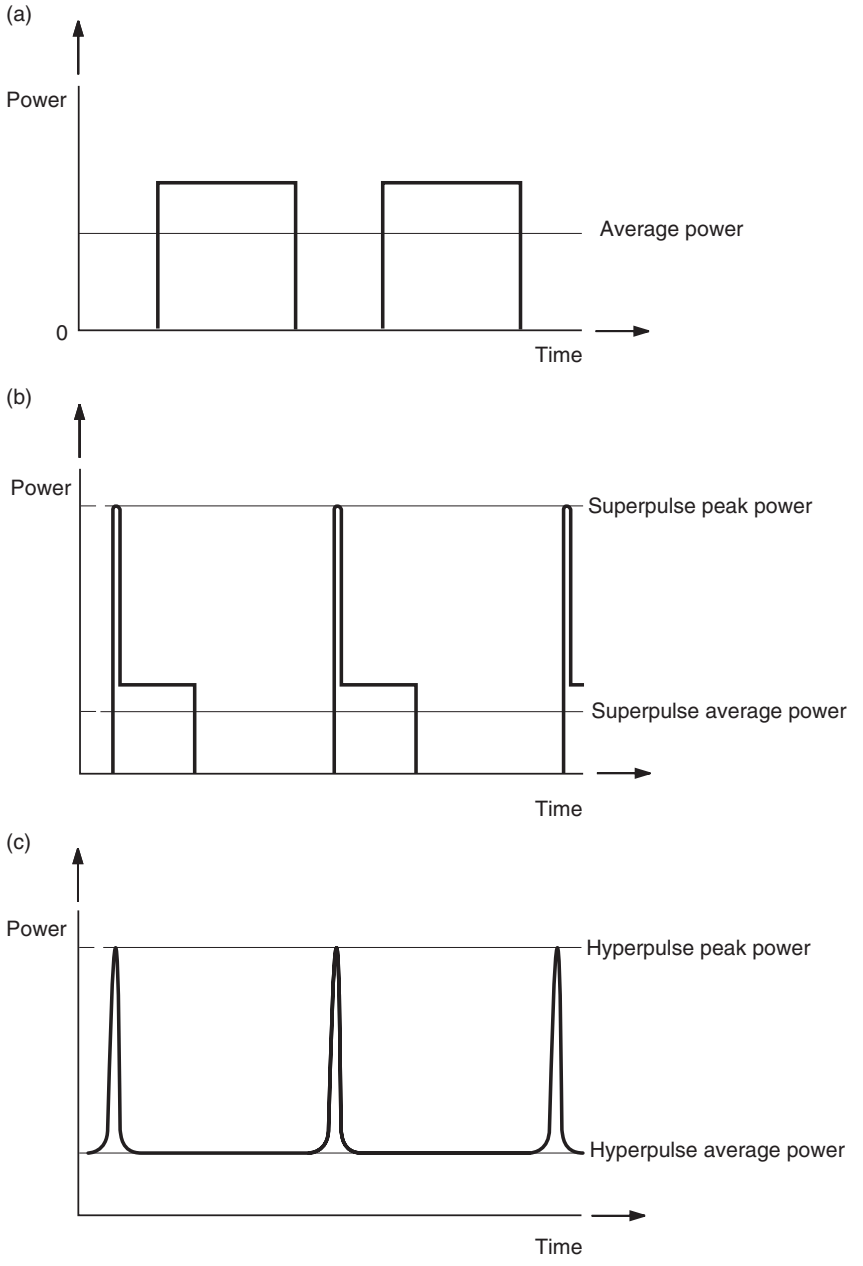


Figure 8.4 Schematic of pulsed outputs. (a) Gated pulsing. (b) Superpulsing. (c) Hyperpulsing.

The spot size (beam radius), $w(x)$, of a propagating diffraction-limited Gaussian beam at any general location along the cavity axis can be expressed as:

$$w(x) = w_{\mathrm{w}}\left[1 + \left(\frac{2x}{L}\right)^2\right]^{1/2} \tag{8.5}$$

where

$$w_w = \left(\frac{L\lambda}{2\pi}\right)^{1/2} \tag{8.6}$$

x is the distance measured from the beam waist along the axis of beam propagation, w_w is the beam radius at the waist, and L is the distance between the cavity mirrors.

The propagation of a non-Gaussian laser beam can be expressed in a similar manner with w_w modified as:

$$w_w = \left(\frac{L\lambda}{2\pi K}\right)^{1/2} \tag{8.7}$$

where K is the K-factor. Both the M^2 and K factors are unity for diffraction-limited Gaussian beams. However, for non-Gaussian beams, the K-factor is smaller (less than unity), while the M^2-factor is higher (greater than unity). A beam with an $M = 3$ has a waist almost three times larger than the Gaussian beam and diverges slightly more. As an example, most high-power Nd:YAG lasers have beam quality of M^2 between 50 and 100.

8.1.6 Beam Absorption

In analyzing the thermal effect of a laser beam on a workpiece, the amount of power that actually enters or is absorbed by the workpiece needs to be considered. When the beam is incident on a solid surface, absorption occurs in a thin surface layer called the skin depth (of about 0.01 μm depth for a CO_2 laser on metals) and transferred to the inner part by conduction. The skin depth, δ', of a substance is the distance to which incident electromagnetic radiation penetrates that substance, during which the light intensity drops by a factor e from its initial value, and is expressed as:

$$\delta' = \frac{1}{\sqrt{\pi \sigma_e \mu_m \nu}} \tag{8.8}$$

where σ_e is the electrical conductivity (S/m), μ_m is the magnetic permeability (H/m), and ν is the frequency (Hz).

For a beam of power q_i, incident on a surface of reflectance R_s, the power q_a absorbed by the workpiece is given by:

$$q_a = q_i(1 - R_s) = a_\alpha q_i \tag{8.9}$$

The reflectance (R_s) or absorptivity ($a_\alpha = 1 - R_s$) depends on both the material on which the beam is incident and the beam wavelength (Figure 8.5a). For metals, the absorptivity increases with decreasing wavelength, while for nonmetals, it increases with increasing wavelength (Figure 8.5a). Metals thus generally tend to have better absorption at the shorter 1.06 μm wavelength characteristic of Nd:YAG lasers than at the 10.6 μm wavelength of CO_2 lasers at temperatures below their melting points. The reverse is true of nonmetallic materials such as glass, quartz, plastics, and insulators in general. For processes that do not require melting, such as surface heat treatment, such differences in absorption can be very significant. Figure 8.5b shows the absorption values for different metals. The absorptivity also increases with an increase in temperature for a given wavelength. Once the material melts, the absorptivity increases dramatically and essentially becomes independent of the wavelength (Figure 8.6).

Initial absorption of the laser beam by the workpiece is also affected by other factors such as the nature of the surface and the joint geometry. Absorption on a metallic surface can be enhanced by appropriate surface modification such as anodizing, sand blasting, and surface coating. However,

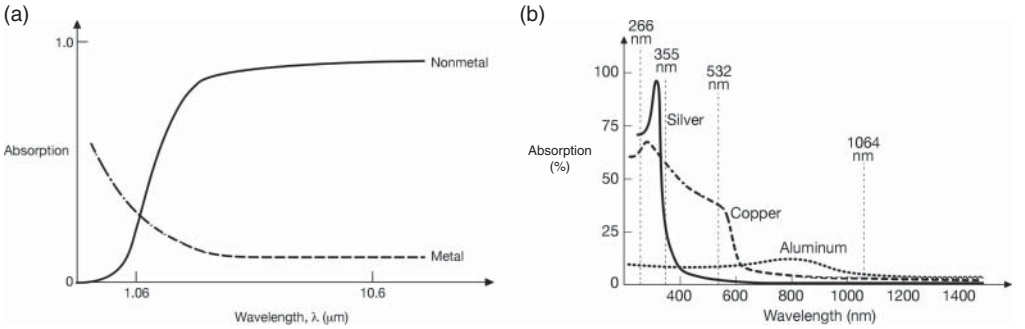


Figure 8.5 (a) Wavelength dependence of absorption for different materials (polished). (b) Wavelength dependence of absorption. Source: Adapted from Büchter (2018).

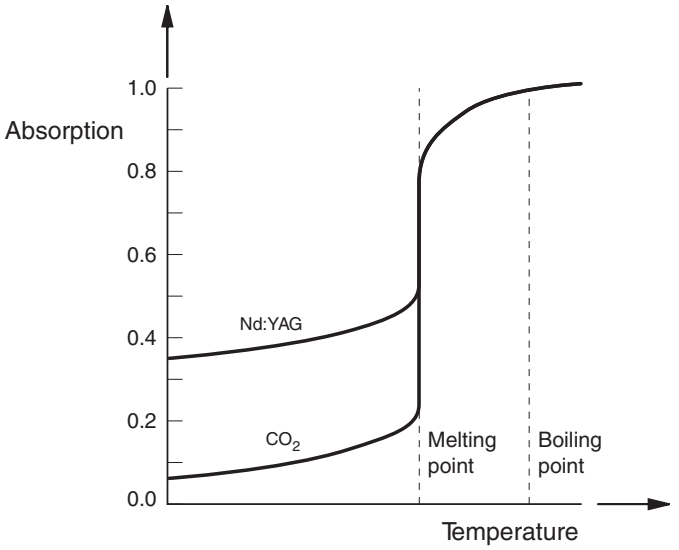


Figure 8.6 Temperature dependence of absorption for Nd:YAG and CO_2 laser beams on metals. Source: From Wilson and Hawkes (1987)/Pearson.

as indicated in the preceding paragraph, the absorptivity of metals increases with increasing temperature and becomes even more significant in the molten state. Thus, for processes that involve melting and/or vaporization, such as welding and cutting, the high initial reflectivity need not be a major problem, except in the initial stages of the process, where a high initial power is required for highly reflective materials. On the other hand, for processes that do not involve melting, such as heat treatment and forming, high reflectivity will result in significant power loss.

In addition to its influence on beam absorption, reflectivity also determines the amount of energy that is reflected back to the beam delivery system (back scattering) and may result in optical feedback into the laser resonator, affecting both the beam generation mechanism and optical elements.

8.1.6.1 Measurement of Absorptivity The absorptivity of a material can be measured using the simple calorimetric setup as shown in Figure 8.7. A specimen of the material is attached to one side of a calorimeter through which water flows. When the laser beam is incident on the specimen, the energy

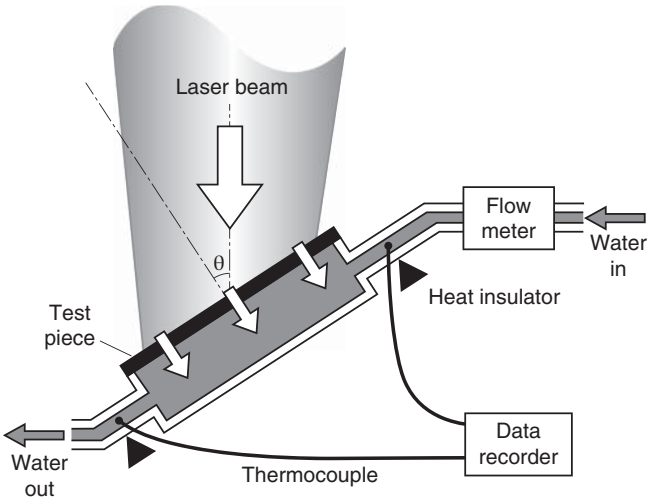


Figure 8.7 Set-up for measuring absorptivity calorimetrically. Source: From *ICALEO 1991 Proceedings*, Vol. 74. Copyright 1992, Laser Institute of America. All rights reserved.

it absorbs is transferred to the water flowing beneath it. Measurement of the inlet and outlet temperatures of the water, along with the flow rate, enables the power absorbed to be determined as follows:

$$\text{Power absorbed by the water} = \dot{m}c_{\text{p}}(T_{\text{out}} - T_{\text{in}}) \tag{8.10}$$

where $\dot{m}$ is the mass flow rate of the water (kg/s), c_{p} is the specific heat (J/kg K), T_{in} is the inlet temperature (K), and T_{out} is the outlet temperature (K).

The absorptivity of the material is then given by:

$$\text{Absorptivity} = \frac{\text{Power absorbed by the water}}{\text{Incident laser power}} \tag{8.11}$$

8.1.7 Beam Alignment

To ensure that the laser beam is delivered from the generator to the appropriate spot on the workpiece, the beam is always aligned with the optical and beam delivery systems. Since the infrared radiation is not visible, the alignment process is normally facilitated by installing in the high-power laser a low-power HeNe laser to generate a visible red beam that is coaxial with the invisible primary infrared beam. With the two beams being coaxial, alignment of the visible beam automatically ensures alignment of the invisible beam. For such systems, the output lens must be transmissive to both infrared and visible radiation. ZnSe is one such suitable material for CO_2 lasers.

The alignment process essentially involves adjustment of the beam bender (mirror) orientation using the mounting screws to redirect the beam. A cross-hair assembly is mounted at the output of either the beam bender or focusing assembly. The mirror orientation is then adjusted until the beam center coincides with the cross-hair center when observed on a screen held behind the cross-hair assembly, at the location where the workpiece would normally be positioned (Figure 8.8).

For systems that do not have a visible beam installed, essentially the same procedure is followed, except that instead of observing the visible radiation on the screen, the laser is given a single pulse at low power, with a sheet of cardboard paper held in place of the screen. The beam will burn a pattern on the paper that indicates the cross-hair position relative to the beam center.

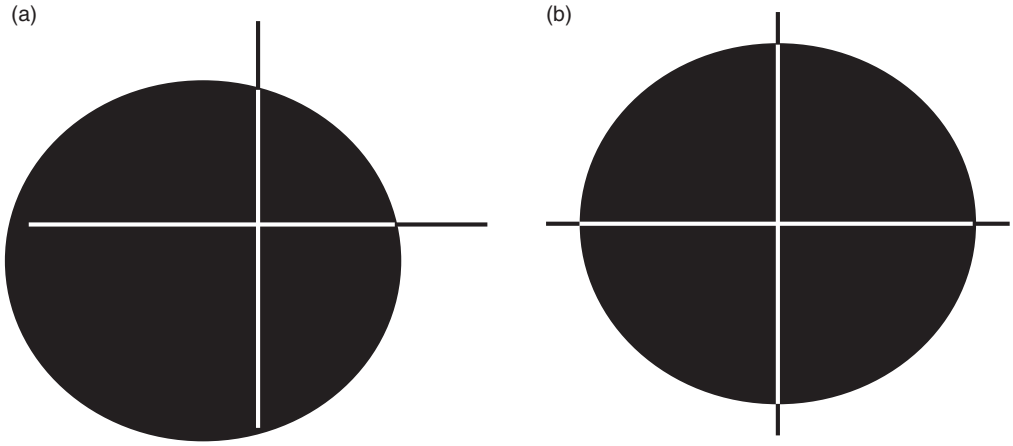


Figure 8.8 Beam alignment using a cross-hair. (a) Misaligned beam. (b) Aligned beam.

8.1.8 Motion Unit

The motion unit is the means by which relative motion is achieved between the laser beam and the workpiece. There are four ways in which this can be done:

1. The workpiece is kept stationary, while the entire laser and its beam delivery system is moved relative to the workpiece. This approach is more suited to large workpieces and is often inconvenient, given the typical size of industrial lasers.
2. The laser and its entire beam delivery system are kept fixed, while the workpiece is moved with respect to it. This method is suitable for processing pieces of limited size. Processing of large pieces requires significantly large floor space.
3. The use of a fiber optic system (Section 3.7.2).
4. The use of flying optics. In this case, the laser generator is kept fixed while the beam delivery components (focusing assembly and mirrors) are moved relative to the workpiece. It usually involves the use of a robot that incorporates the beam delivery elements. The principal setback of this approach is the change in size of the focused spot along the processed path. This is due to beam divergence which results in increase in beam size with distance from the laser output window. Minimal beam divergence is thus essential.

Either the fiber optic or flying optics arrangement minimizes floor space, and the size of workpiece that can be processed is unlimited. Another advantage is the ability to keep the beam orientation normal to the workpiece surface, and to keep the nozzle to workpiece distance constant, even when processing parts with complex contours.

8.2 PROCESS EFFICIENCY

Laser processing applications involve transfer of energy from the generator to the workpiece. In the course of generation and transfer, part of the energy is lost. Sources of energy loss include

1. Absorption in the optical components.
2. Convection through the shielding gas.

3. Radiation from the workpiece surface to the surroundings.
4. Conduction into the workpiece.

Thus, only a fraction of the power generated is available for processing. The overall efficiency of a laser process consists of four efficiency components that determine the amount of power that is actually used in the process. These are as follows:

1. Laser generator efficiency.
2. Beam transmission efficiency.
3. Heat transfer efficiency.
4. Processing efficiency.

- *Laser Generator Efficiency*, η_g: This is given by the ratio of the power output of the laser generator to the power input to the generator. It is expressed as:

$$\eta_g = \frac{\text{Power output of the laser}}{\text{Power input to the laser generator}} \tag{8.12}$$

It typically ranges from about 10% to 30% for CO_2 lasers, 1% to 3% for Nd:YAG lasers, 1% to 4% for excimer lasers, 30% to 50% for diode lasers, and 25% to 50% for fiber lasers (Appendix 2.B).

- *Beam Transmission Efficiency*, η_t: The beam transmission efficiency depends on the transmission components being used. The efficiency is defined as the ratio of the energy coming out of the beam delivery system to the energy input to the beam delivery system from the laser and is expressed as:

$$\eta_t = \frac{\text{Power output from beam delivery system}}{\text{Power input from laser to beam delivery system (power output of the laser)}} \tag{8.13}$$

- *Heat Transfer Efficiency*, η_{th}: The actual power available for the process is determined by the efficiency of heat transfer, η_{th}, which is the efficiency of energy transmission from the source to the workpiece (absorption), that is,

$$\eta_{th} = \frac{\text{Power actually absorbed by the workpiece } (q_a)}{\text{Total power available at the source (power output from beam delivery system) } (q_i)} \tag{8.14}$$

Thus,

$$q_a = \eta_{th} \times q_i \tag{8.15}$$

For laser welding, this typically ranges from 50% to 95%, once the welding process is initiated and a keyhole is formed. However, before any melting occurs, this may be as low as 1% for polished aluminum or copper and about 40% for stainless steel when using a CO_2 laser. Nd:YAG lasers result in higher absorption efficiencies in solid metals. Power losses at this stage result primarily from reflection from the workpiece surface and plasma absorption when the material melts.

- *Processing Efficiency*, η_p: The processing efficiency is the thermal efficiency of energy conversion in the workpiece and is defined as the ratio of the power required for heating and/or melting to the power actually transferred to the workpiece.

For processes that involve melting, the processing efficiency, η_p, is given by:

$$\eta_p = \frac{\text{Power required to melt a given volume of material}}{\text{Power actually absorbed by the workpiece}} = \frac{Q_1 A_w u_x}{q_a} \tag{8.16}$$

where $Q_1 = \rho_w [L_m + c_p(T_m - T_0)]$ (J/m^3) is the heat required to melt a given volume of material from room temperature, A_w is the total cross-sectional area of fusion zone (m^2), ρ_w is thedensity of fused metal (kg/m^3), and u_x is the velocity of the moving heat source (m/s).

The processing efficiency depends on the processing conditions, material (workpiece), and joint configuration. Materials with higher conductivity have lower efficiency, since a significant amount of the input heat is lost by conduction. It increases with the intensity of the heat source. Higher efficiencies are achieved at higher power levels and speeds for lower thermal conductivity materials, since less incident power is then lost by conduction into the workpiece. Power is also lost by radiation and convection from the surface.

- *Overall Efficiency*, η: The overall efficiency of the process is then given by:

$$\eta = \eta_g \eta_t \eta_{th} \eta_p = \frac{\text{Power used in processing the material}}{\text{Power input to the laser generator}} \tag{8.17}$$

Typical values of overall efficiency for arc, laser, and electron beam welding are shown in Table 10.3 for comparison.

8.3 DISTURBANCES THAT AFFECT PROCESS QUALITY

High quality in the production environment can be maintained by minimizing the impact of disturbances that affect the production system. Common sources of disturbance during laser processing include

1. Fluctuations in power line.
2. Temperature and humidity variations in the production environment.
3. Condensation that may result from cooling of optical components with tap water. Air cooling alleviates this problem.
4. Contamination of the output window by dust and vapor from the atmosphere. This requires prevention of contact with surrounding atmosphere.
5. Vibrations from the surroundings.
6. Back scattering, a phenomenon that results from reflection of the incident laser beam off the workpiece. If the nozzle orifice is large enough, a significant portion of this reflected beam may propagate backward along the transmission path. In extreme cases, this may change the state of polarization of the laser beam, cause uncontrolled amplification of the laser process, and possibly damage the optical elements and laser resonator. The effect of back scattering on the process may be minimized by tilting the workpiece at such an angle that any beam that is reflected from it is unable to propagate back into the beam delivery system.

8.4 GENERAL ADVANTAGES AND DISADVANTAGES OF LASER PROCESSING

Lasers have several advantages over conventional techniques that are normally used in materials processing. At the same time, there are disadvantages that may make lasers less suitable for certain applications. These are outlined in the following paragraphs and further elaborated on with respect to specific processes as they are discussed in Chapters 9–14.

8.4.1 Advantages

1. The ability to focus the beam to a small size, thereby attaining a high intensity and highly localized heating source with minimal effect on surrounding areas. Spot sizes may vary from about 0.001 to 10 mm.
2. The ease with which the beam power can be controlled by regulating the current through the electric discharge.
3. Minimal contamination of the process.
4. Ability to manipulate the beam into ordinarily inaccessible areas using mirrors and fiberoptic cables.
5. Minimal heat-affected zone and distortion.
6. Noncontact nature of the process.

8.4.2 Disadvantages

1. Relatively high capital cost of equipment.
2. High reflectivity of laser beam on metals.
3. Low efficiency of lasers.
4. Energy waste by beam dumping when the laser is not used continuously.

8.5 SUMMARY

Power densities generated by pulsed lasers generally exceed 10^{14} W/cm^2 and can be as high as 10^{21} W/cm^2, much greater than the outputs of commonly available heat sources such as the electron beam, arc, and blackbody radiation. For materials processing, ablation requires the highest power densities and shortest interaction times, while transformation hardening requires the lowest power densities and longest interaction times. The principal system parameters that affect process quality include the beam power and power density at the focal point, wavelength and focusing system, beam mode (distribution), beam form (CW or pulsed), and beam quality. Others include beam absorption, beam delivery, beam alignment, and the motion unit used for positioning the beam or workpiece.

The power density of a beam at the focal point is inversely proportional to the square of the beam wavelength and lens focal length and directly proportional to the square of the original beam diameter. Common lens materials that are used for focusing include ZnSe, KCl, GaAs, and CdTe. The depth of focus of a focusing system is the distance over which the workpiece can be moved from the original focal point without a significant change in the focused beam radius.

The Gaussian mode is the lowest order mode a beam can have, and it can be focused to the laser's theoretical minimum radius. The size of a Gaussian beam is defined as the radial distance at which the beam intensity reduces to $1/e^2$ or 13.5% the peak intensity. The higher-order or multimode beams, on the other hand, are more spread out and are not as well focused. The form of the laser output can be either continuous or pulsed, with the pulsed output being one of three types: gated, superpulsed, or hyperpulsed. For a given total pulse energy, a shorter but higher power density pulse will result in higher peak surface temperature than a longer, lower power density pulse. However, the lower power density pulse produces a more uniform temperature distribution and a greater depth of penetration of the heat at the end of the pulse. The quality of a laser beam is characterized by either the M^2 or K $(=1/M^2)$ factor, with $M = K = 1$ for a diffraction-limited Gaussian beam.

When a beam is incident on a solid surface, absorption occurs in a thin surface layer, the skin depth, and transferred to the inner part by conduction. For metals, the absorptivity increases with decreasing wavelength, while for nonmetals, it increases with increasing wavelength. The absorptivity also increases with an increase in temperature for a given wavelength.

As the laser beam propagates from the generator to the workpiece, the major sources of energy loss include absorption in the optical components, convection through the shielding gas, radiation from the workpiece surface to the surroundings, and conduction into the workpiece. The overall efficiency of a laser process consists of four efficiency components. These are laser generator efficiency, beam transmission efficiency, heat transfer efficiency, and processing efficiency.

APPENDIX 8.A

List of symbols used in the chapter

Symbol	Parameter	Units
a_α	Absorptivity	—
A_w	Cross-sectional area of fusion zone	m^2
d_f	Focused beam diameter	m
q_a	Absorbed power	W
q_i	Incident power	W
Q_1	Heat required to melt a given volume of material from room temperature	J/mm^3
R_s	Reflection coefficient of surface	—
T_{in}	Inlet temperature	K
T_{out}	Outlet temperature	K
z_f	Depth of focus	m
δ'	Skin depth	μm
η_g	Generator efficiency	%
η_p	Process efficiency	%
η_t	Beam transmission efficiency	%
η_{th}	Heat transfer efficiency	%
ρ_w	Density of fused metal	kg/mm^3

PROBLEMS

8.1. How would you expect the processing efficiency to vary with the thermal conductivity of the base metal?

8.2. Using Eqs. (1.29) and (1.30), show that the depth of focus, z_f, over which 80% of the power density at the focal point is maintained is given approximately by:

$$z_f = \frac{\pi \times {d_f}^2}{4 \times \lambda}$$

where d_f = focused beam diameter.

8.3. A CO_2 laser beam with a Gaussian mode is focused using a 150 mm lens. If the beam radius as it enters the lens is 12.5 mm, determine:
(a) The focused spot radius.
(b) The depth of focus for a spot radius increase of 7.5%.

8.4. (a) Two laser beams of the same power are incident on a workpiece. One is from a Nd:YAG laser and the other from a CO_2 laser. Which one would result in the most intense heat at the point of focus? Explain.
(b) The maximum power of a laser available in a plant is found to be inadequate to weld a new material that requires joining. Then a self-styled consultant (qualifications unknown) suggests that the beam diameter should be increased before it gets to the focusing lens. Is this a knowledgeable or quack consultant? Explain.

8.5. A given material has the same absorption coefficient for both CO_2 and Nd:YAG lasers. Disregarding cost, which of the two lasers would you select for welding the material? Explain.

8.6. A 4 kW CO_2 laser is used for a process that requires a power density of 5×10^7 W/cm^2 at the focal point. If the input beam diameter is 20 mm and has a Gaussian profile,
(a) What should the focal length of the lens be?
(b) What should the lens focal length be if a non-Gaussian beam with a K-factor of 0.7 is used instead?

8.7. Consider a laser beam of initial radius w_0, which is focused to a radius w_f. using a lens of focal length, f_l. Let the laser wavelength be λ, the input power be q, and the average power density be PD.
(a) Determine the percentage change in initial beam radius necessary for a $\pm 25\%$ change in power density at the focal point.
(b) Calculate the initial beam radius necessary to obtain a 4.5×10^6 W/mm^2 power density at the focal point for a 6 kW CO_2 laser focused using a 125 mm lens.

8.8. A CO_2 laser beam with a Gaussian mode is focused using a 150 mm lens. If the beam size as it enters the lens is 12.5 mm, determine:
(a) The focused spot size.
(b) The depth of focus for a spot size increase of 7.5%.

8.9. Consider a laser beam focused through a lens.
(a) Calculate how much the initial beam radius should be expanded or shrunk for a change in focus intensity of $\pm 20\%$.
(b) If the output intensity of a 4.5 kW Nd:YAG laser focused through a 150 mm lens is 3.5×10^8 W/mm^2, what is the initial beam radius?
(c) As a consultant, you must build a beam expander that can adjust the focus intensity as detailed in (a). Calculate what range of focal lengths of lenses are needed if the first lens in the beam expander has a focal length of 500 mm.
(d) A collimated beam before the final focusing lens is critical for proper processing. A diverging beam makes the focus location and power density unpredictable. In order to

have a collimated beam after the beam expander, the two lenses must be placed at a distance equal to the sum of their focal lengths. If the lenses need 25 mm of clearance on either side, how much space must be allowed for this beam expander?

(e) What would happen to the beam expander if the laser was changed to a CO_2 laser? Assume all other values as before.

8.10. What will be the M^2 factor for a non-Gaussian Nd:YAG laser beam which has a beam waist of 0.965 mm and beam radius of 1.04 mm at a distance of 2 cm from the waist? What effect will this difference have on using this laser for drilling?

BIBLIOGRAPHY

Büchter, E. (2018). Laser cleaning is cost-effective and reliable. *Industrial Laser Solutions* 16–17.

Gregersen, O. and Olsen, F.O. (1990). Beam analyzing system for CO_2 lasers. *Proceedings of the 7th International Conference on Applications of Lasers and Electro Optics, ICALEO'90*, Boston, MA, USA, Laser Institute of America, pp. 28–35.

Gutierrez-Miravete, E. (1993). Mathematical modeling of rapid solidification. In: *Rapid Solidification Technology: An Engineering Guide* (ed. T.S. Srivatsan and T.S. Sudarshan), 3–70. Lancaster: Technomic Publishing Co. ISBN: 0-87762-926-9.

Harry, J. (1974). *Industrial Lasers and Their Applications*. London: McGraw Hill.

Lewis, G.K. and Dixon, R.D. (1985). Plasma monitoring of laser beam welds. *Welding Journal* 64: 49s–54s.

Matsunawa, A. (1990). Physical phenomena and their interpretation in laser materials processing. *Proceedings of the 7th International Conference on Applications of Lasers and Electro Optics, ICALEO'90*, Boston, MA, USA, Laser Institute of America, p. 313.

Mazumder, J. (1987). An overview of melt dynamics in laser processing. *SPIE, High Power Lasers* 801: 228–241.

Ready, J.F. (1965). Effects due to absorption of laser radiation. *Journal of Applied Physics* 36 (2): 462–468.

Rofin-Sinar *Technical Note – CO_2 Laser Welding*. Plymouth, Michigan: Rofin Sinar.

Shibata, K., Sakamoto, H., and Matsuyama, H. (1991). Absorptivity of polarized beam during laser hardening. *Proceedings of the 7th International Conference on Applications of Lasers and Electro Optics, ICALEO'90*, Laser Institute of America, pp. 409–413.

Schuocker, D. (1986). Laser cutting. *SPIE, The Industrial Laser Annual Handbook* 629: 87–107.

Thompson, A. (1986). CO_2 laser cutting of highly reflective materials. *SPIE, The Industrial Laser Annual Handbook* 629: 149–153.

von Allmen, M. and Blatter, A. (1994). *Laser-Beam Interactions with Materials*. Berlin: Springer Verlag.

Watson, M.N., Oakley, P.J., and Dawes, C.J. (1985). Laser welding: techniques and testing. *Metal Construction* 17: 25–28.

Wilson, J. and Hawkes, J.F.B. (1987). *Lasers: Principles and Applications*. New York: Prentice Hall International Series in Optoelectronics.

9 Laser Cutting and Drilling

In laser cutting and drilling, the focused laser beam is directed unto the surface of the workpiece to rapidly heat it up, resulting in melting and/or vaporization, depending on the beam intensity and workpiece material (Figure 9.1). The molten metal and/or vapor is then blown away using an assist gas. The power density required is typically of the order of 10^6–10^7 W/cm^2 for metals. Lasers can be used to effectively cut metal plates of thicknesses up to about 10 cm. The cut surfaces are roughly parallel and straight-edged. In reality, the diverging nature of the beam results in a slight taper being produced. The workpiece thickness that can be cut with parallel sides is determined by the depth of focus (Section 8.1.2). Plates that are thicker than the depth of focus normally result in tapered surfaces. Furthermore, for effective removal of material from the cutting zone, the ratio of workpiece thickness to kerf width needs to be less than 20 : 1 for most metals, and less than 40 : 1 for ceramics. The kerf is the slot or opening created by the cutting process.

The total heat input required for laser cutting is relatively small. This results in a small heat-affected zone (HAZ) size, of the order of 0.1 mm. Also, the small size of the focused beam results in very narrow kerf sizes, typically about 0.05–1 mm. Laser cutting is used for both straight and contour cutting of sheet and plate stock in a wide variety of materials.

Laser cutting and laser drilling are discussed in two separate sections due to the unique characteristics of the two processes. The chapter as a whole begins with a general discussion on the different ways in which laser cutting may take place. Components of a laser cutting system are then outlined. The processing conditions that are normally used are then presented. This is followed by a discussion on the principles of the cutting process, which leads to common defects that may result during the process. The specific characteristics of individual materials are then discussed, followed by the advantages and disadvantages of laser cutting, and a comparison with conventional processes. The section on laser drilling follows a similar outline, starting with the different forms of laser drilling. The essential process parameters are then discussed, followed by an analysis of the material removal process. The advantages and disadvantages of the process are then presented, followed by some of the major applications of laser drilling. Finally, new developments such as micromachining and laser-assisted machining are discussed.

9.1 LASER CUTTING

9.1.1 Forms of Laser Cutting

The process of laser cutting can occur in one of three forms:

1. Fusion cutting.
2. Sublimation cutting.
3. Photochemical ablation.

Principles of Laser Materials Processing: Developments and Applications, Second Edition. Elijah Kannatey-Asibu, Jr.
© 2023 John Wiley & Sons, Inc. Published 2023 by John Wiley & Sons, Inc.
Companion Website: www.wiley.com/go/Kannatey-Asibu/PrinicplesLaserMaterialsProcessing

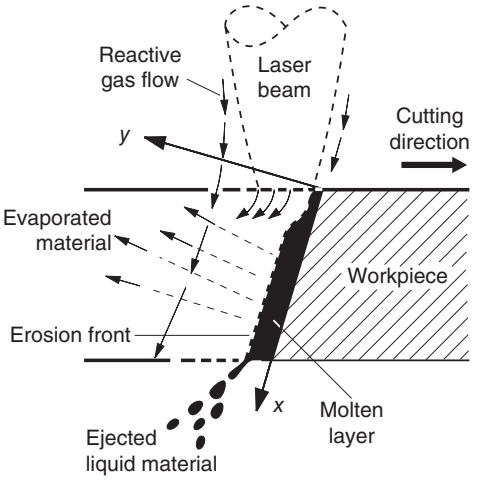


Figure 9.1 Schematic of the laser cutting process. Source: From Ready (2001)/Laser Institute of America.

9.1.1.1 Fusion Cutting Fusion cutting involves melting of the base material, which is then ejected using a high-pressure assist gas. The assist gas may be an inert gas, in which case the energy for melting is provided entirely by the laser beam. It may also be oxygen (or air), which reacts with the base metal, and the resulting exothermic reaction provides additional energy to enhance the process. The term fusion or clean cutting is sometimes used to indicate inert gas-assisted cutting, while the process involving exothermic reaction is then referred to as gas cutting. A major problem of fusion cutting is the formation of striations (valleys and peaks that run along the thickness (Section 9.1.5.1)) on the cut surface and dross (molten material that clings to and solidifies on the underside of the cut edge as burr (Section 9.1.5.2)) at the lower cut edge. However, the fusion cutting process is more efficient, requiring less energy per unit volume of material removed as compared to the other methods.

9.1.1.2 Sublimation Cutting In sublimation cutting, the workpiece material is vaporized along the cutting seam. This is often achieved using a pulsed beam, and a jet of inert assist gas that is coaxial with the beam is used to blow away the vapor produced. It is limited to thin sections since more energy is required to remove a unit volume of material as compared to fusion cutting. However, it has the advantage of a narrower kerf width and higher-quality surface. Pulsed beams with high peak power may be necessary when surface quality is critical.

9.1.1.3 Photochemical Ablation Organic materials, ceramics, or difficult-to-cut materials in general are normally cut by this method or by sublimation cutting. Organic compounds tend to absorb ultraviolet radiation in an efficient manner. The photon energy levels of lasers based on ultraviolet radiation range between 3.5 and 6.5 eV. This corresponds with the energy levels required for molecular bonding. For example, the energy associated with the C—C bond is roughly 4.6 eV, while that of the C—H bond is about 4.2 eV. Compare these values with the photon energy level values associated with CO_2 and Nd:YAG lasers of about 0.12 and 1.2 eV, respectively.

As a result, when an organic material is irradiated with an ultraviolet beam, the material absorbs the beam's energy in a very thin layer near the surface, of the order of submicrons. This breaks molecular bonds, causing ablative decomposition of the irradiated area. The process occurs almost

instantaneously (about 20 ns duration), and since the thermal conductivity of organic materials is relatively low, the resulting edges are well defined, with minimal thermal damage to the surrounding area. Thus, the cut region is cleaner and smoother compared to that obtained using CO_2 and Nd:YAG lasers. The process is sometimes referred to as cold cutting since little heat is generated.

9.1.2 Components of a Laser Cutting System

The basic components of a cutting system are illustrated in Figure 9.2 and include

1. The laser generator that produces the beam.
2. A beam delivery system for directing the beam to the workpiece.
3. A nozzle assembly, usually integral with the focusing assembly and coaxial with the beam, for directing the assist gas to the workpiece.
4. A motion unit for providing relative motion between the laser beam and the workpiece.
5. An exhaust for the waste material.

The beam normally emerges from the generator horizontally and is deflected vertically downward by a bending mirror. The vertical orientation is used to minimize trapping of the beam and molten material blown through the cut. The beam is then focused by the lens unto the workpiece. At the same time, a gas jet is directed through a nozzle attached to the tip of the focusing assembly, unto the workpiece. A typical nozzle diameter would be about 1–2 mm. The delivery pressures are normally maintained at about 3–4 bar (45–60 psi or 0.3–0.4 MPa) in the gas jet nozzle for cutting thin materials at high speeds. At high gas pressures, it is often necessary to use relatively thick lenses that can withstand the pressure. However, care has to be taken since thermal deformation of the lens increases with its thickness. Thus, at high power levels, say 10 kW, it is preferable to use reflective optics (mirrors) rather than transmissive optics (lenses).

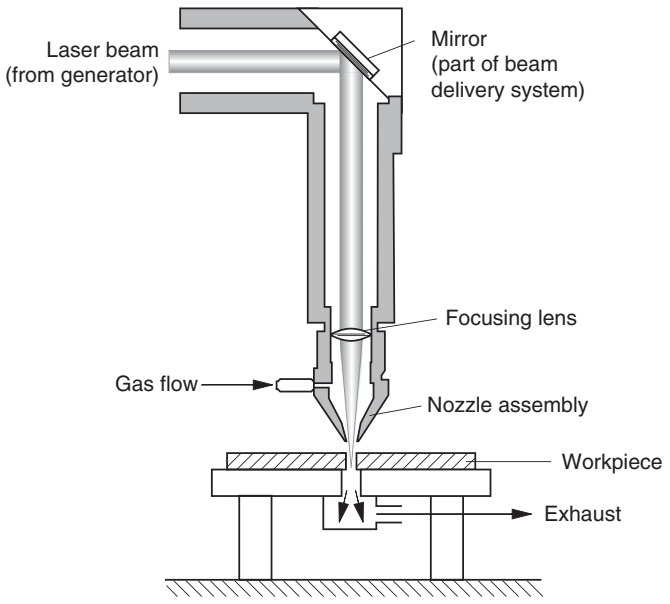


Figure 9.2 Components of a laser cutting system.

Table 9.1 Sample cutting conditions for laser cutting of different materials.

Material	Thickness (mm)	Power (W)	Cutting speed (m/min)	Gas pressure (kPa)	Gas type	Laser
Carbon steel	0.5	250	3.5			CO_2
Carbon steel	6.0	1200	1.5			CO_2
Carbon steel	4.0	2000	1.2	150	O_2	Fiber
Carbon steel	100.0	6000	5×10^{-3}	1000	Air	Fiber
Stainless steel	100.0	6000	7×10^{-3}	1000	Air	Fiber
Stainless steel	1.0	1000	3.5	600	N_2	CO_2
Stainless steel	6.0	2000	1.0	1200	N_2	CO_2
Aluminum	1.0	1200	3.0	600	N_2	CO_2
Aluminum	3.0	1800	1.0	1000	N_2	CO_2
Titanium	1.0	800	3.5	600	Ar	CO_2
Titanium	1.5	900	3.0	700	Ar	CO_2

Source: Adaptef from Ready (2001); Shin et al. (2018).

The beam is generally oriented in a direction almost normal to the workpiece surface, and there is a deterioration in product quality as the beam orientation deviates significantly from the normal direction. The distance from the nozzle tip to the workpiece surface is typically maintained constant at about 0.3 mm to minimize expansion of the gas flow. Capacitive or light sensors may be installed to measure this distance. The focusing unit may then be moved using a computer numerically controlled (CNC) third axis or a robot. A simpler but less flexible approach involves mounting the focusing unit on rollers that glide on the workpiece surface.

Directly opposite the nozzle, on the other side of the workpiece, an exhaust system is provided to absorb the transmitted beam, molten debris from the cut, and exhaust vapors. This normally consists of a vacuum pump that draws the exhaust to a disposal unit.

9.1.3 Processing Conditions

The principal parameters that affect the laser cutting process include the following:

1. Beam power.
2. Beam characteristics.
3. Traverse speed.
4. Assist gas type and flow.
5. Location of focal point relative to the workpiece surface.

Table 9.1 gives sample conditions for laser cutting of various materials. We shall further discuss each of these parameters in Sections 9.1.3.1–9.1.3.5.

9.1.3.1 Beam Power The power is the most significant of the parameters listed. An increase in power increases the maximum thickness that can be cut and/or speed at which it can be cut.

9.1.3.2 Beam Characteristics The most important beam characteristics in laser cutting operations are as follows:

1. Beam mode.
2. Stability.

3. Polarization.
4. Beam form (pulsed or continuous wave).

Beam Mode As indicated in Section 8.1.3, the beam mode is an indication of how the energy intensity is distributed over the beam cross section. The mode relates to the beam's ability to be focused and may be compared with the sharpness of a cutting tool. In laser cutting, it is desirable to have the beam distribution as close as possible to the fundamental or Gaussian-distributed TEM_{00} mode. This is the mode that can be focused to the laser's theoretically smallest possible focal size and thus the highest density for a given power. This reduces the kerf width and increases cutting speeds and thickness of materials that can be cut. Since higher-order or multimode beams are more spread out (Figure 8.3), they result in larger focal spot sizes and thus lower power density for the same output power.

Beam Stability Stability of the beam is necessary to ensure that the beam power, mode, and direction (pointing stability) remain constant with time. An unstable beam affects the tolerances and surface finish achievable with laser cutting. A stable beam thus reduces variations in product output and enhances quality. The stability of the beam is a characteristic of the laser generator and depends on its design.

Effect of Beam Polarization The basic concepts of polarization are discussed in Section 3.4. When the output of a laser beam is randomly polarized, it shows up in the quality of the cut workpiece as variations in kerf, edge smoothness, and perpendicularity. This is mainly due to the impact of polarization on beam absorption by the material. Absorption of the laser beam depends on both the angle of incidence and polarization. Experimental results indicate that in CO_2 laser cutting, beams which are plane-polarized in a plane parallel to the cutting direction result in cutting speeds which can be up to 50% greater than those which are plane-polarized in the plane normal to the cutting direction. Generally, the circularly polarized beam results in higher cutting speeds at higher power levels. For beams which are plane-polarized in the cutting plane, the beam is found to be more effectively absorbed at angles of incidence greater than 80° and less than 90°.

The cut quality varies with the orientation of the plane of polarization, and a circularly polarized beam generally results in better cut quality when the cutting direction changes. Thus, profile or contour cutting is preferably done using a circularly polarized beam. The impact of the orientation of a linearly polarized beam on the cut quality is illustrated schematically in Figure 9.3. As the figure shows, when the beam is polarized in the cutting direction, the resulting cut may have a narrow kerf with sharp, straight edges. However, as the plane of polarization is oriented away from the cutting direction, the energy absorption decreases, and as a result, the cutting speed is reduced, the kerf becomes wider, and the edges rougher and not square to the material's surface.

Beam Form Both pulsed and continuous wave (CW) beams can be used for laser cutting, with CW beams being more common. For pulsed beams, the quality of cut is affected by the pulse duty, with the surface roughness decreasing with increasing pulse duty. The pulse duty is the ratio of the length of time during which the laser beam is on in one cycle to the total length of the cycle time. In addition, the maximum cutting speeds that can be achieved are significantly reduced at low pulse repetition rates (number of pulses per second). This is due to the fact that at low pulse rates, there is enough time between pulses for the material to substantially cool down. This helps extinguish the exothermic oxidation reaction, thereby reducing the overall process efficiency. Furthermore, as the material cools down between pulses at low pulse frequencies, there is a greater likelihood of forming dross (Section 9.1.5.2). The resulting lower average temperature increases

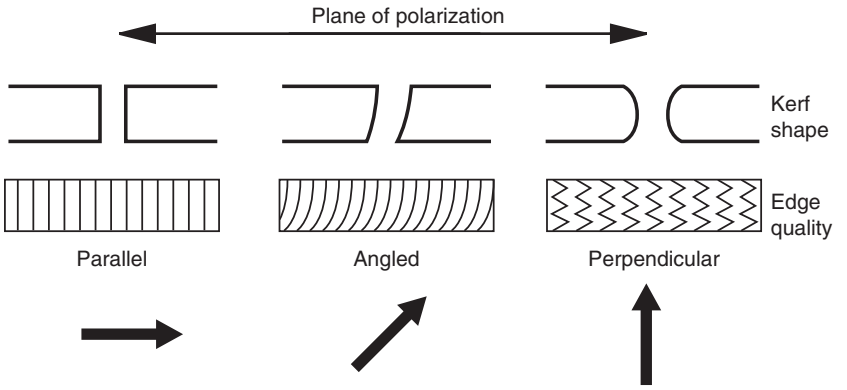


Figure 9.3 Effect of plane-polarized laser orientation on cut quality.

the surface tension or viscosity of the molten material, making it more difficult to flow out of the reaction zone.

As discussed in Section 8.1.4, superpulsing and hyperpulsing can also be used to obtain significant improvements in process performance such as

1. Minimizing dross formation at the lower surface during cutting of some materials.
2. Greater cut thickness (up to about 100%) to be achieved for the same average power as a CW output.
3. Increased processing speeds (20–33%) for the same average power as a CW output.
4. Making it easier to process highly reflective and/or high thermal conductivity materials.

9.1.3.3 Traverse Speed The maximum speed that can be achieved for a given laser power decreases with increasing thickness of the workpiece (Figure 9.4). For a given power, a plot of the variation of cutting speed with thickness will generally have two limiting curves (Figure 9.5). The upper curve indicates the maximum speed that can be achieved for a given thickness. Above this curve, cutting is incomplete. Below the lower curve, self-burning occurs. In other words, the material continues to burn without the aid of the laser. This often widens the kerf and produces a rough surface.

9.1.3.4 Assist Gas Functions The assist gas used in laser cutting serves one or more of the following functions:

1. Facilitates ejection of the molten metal through the backside of the workpiece.
2. Protects the lens from spatter.
3. Acts as a heat source where it results in an exothermic reaction that aids in cutting, such as may occur in oxygen-assisted cutting of steel.

Effect of Different Types of Assist Gases Common gases used include oxygen, inert gases, nitrogen, and air. Oxygen or air is used for exothermic reaction while cutting and therefore they improve the cutting efficiency. Otherwise, an inert gas (usually argon) is used to assist in ejecting the molten metal without oxidation. In cases where the exothermic reaction is desired, the decision to use oxygen or air depends primarily on economics. Air is cheaper but would require higher flow rates for the same

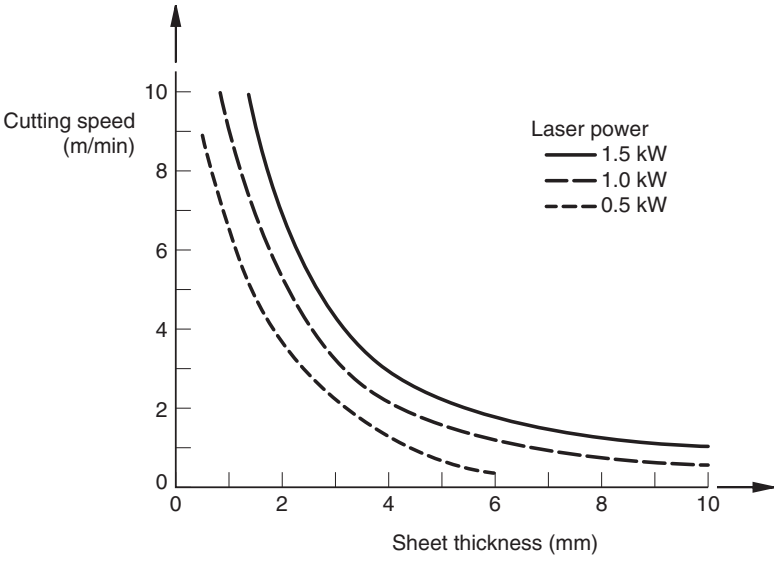


Figure 9.4 Variation of cutting speed (maximum cutting rates) with workpiece thickness in oxygen-assisted laser cutting of steel plate.

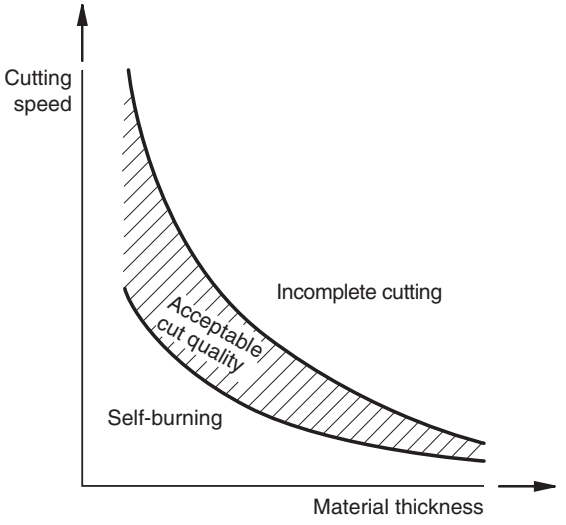


Figure 9.5 Limiting curves for cutting speed with workpiece thickness variation in laser cutting.

amount of thermal energy produced. Furthermore, it may introduce other gases such as nitrogen into the cut surface, making it more brittle.

One setback of oxygen-assisted cutting is the deposition of an oxide layer on the cut surface, giving it a dark appearance. Depending on the subsequent use of the cut parts, it may be necessary to clean off this oxide layer (by grinding or wire brushing). The use of inert gases or nitrogen eliminates the formation of the oxide layer. However, this may significantly reduce the cutting speeds that can be achieved. Furthermore, higher pressures are then necessary to reduce dross formation. Dross is molten material that clings to and solidifies on the underside of the cut edge as burr (Section 9.1.5.2).

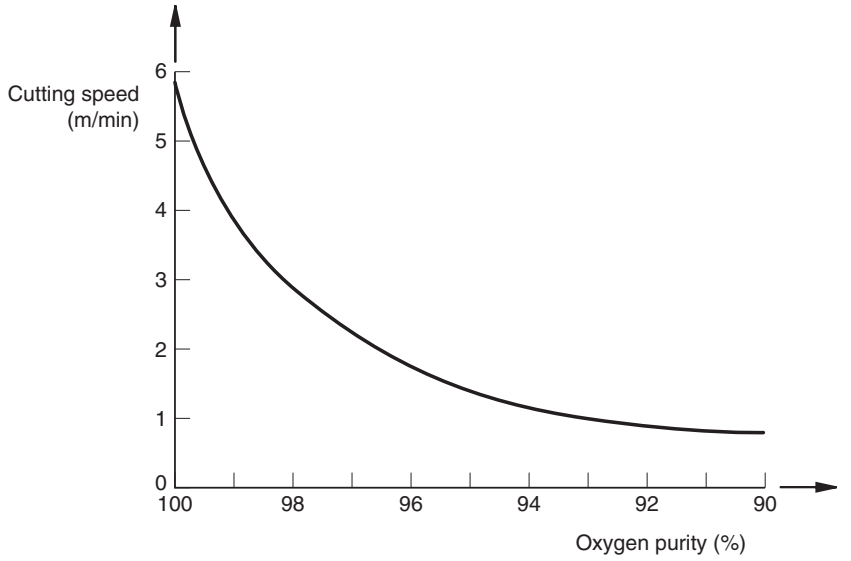


Figure 9.6 Cutting speed as a function of oxygen gas purity. Mild steel of thickness 2 mm, which is cut using 800 W at 2.5 bar pressure. Source: From *ICALEO 1992 Proceedings*, Vol. 75. Copyright 1993, Laser Institute of America. All rights reserved.

Small levels of impurity can cause significant deviations in cutting performance (such as a reduction in maximum cutting speed or increase in dross adhesion) compared to that of the pure gas, be it oxygen or inert gas. This is illustrated in Figure 9.6, which shows a variation of the cutting speed for various levels of oxygen gas purity. This sensitivity to contamination is due to the build-up of a boundary layer of the contaminant at the liquid-cut front interface. It decreases the oxidation rate of the material, thereby lowering the energy input to the cut zone.

Oxygen-Assisted Cutting To summarize the impact of oxygen assist in cutting of various materials, Figure 9.7 illustrates the maximum cutting speeds achievable for different materials when cut using inert gas and also oxygen assist. In addition to the exothermic reaction that results, the use of oxygen assist in cutting also reduces the viscosity and surface tension for some metals, making it easier for the molten metal to flow. Table 9.2 compares the surface tension of some metals and oxides. The oxide film formed also tends to increase beam absorption.

The maximum cutting speeds achieved depend on the thermal properties of the metal. For inert gas-assisted cutting, higher speeds are obtained for low melting and low thermal conductivity metals. Maximum cutting speeds are higher for oxygen-assisted cutting, compared to inert gas-assisted

Table 9.2 Surface tension values for some metals and oxides.

Material	Melting temperature (°C)	Surface tension (dyne/cm)
Cu	1085	1350
Fe	1538	1700–1800
Al_2O_3	2050–2400	360–570
Cr_2O_3	2350–2500	810
FeO		580

Source: Table 1, Data from Miyamoto and Maruo (1991).

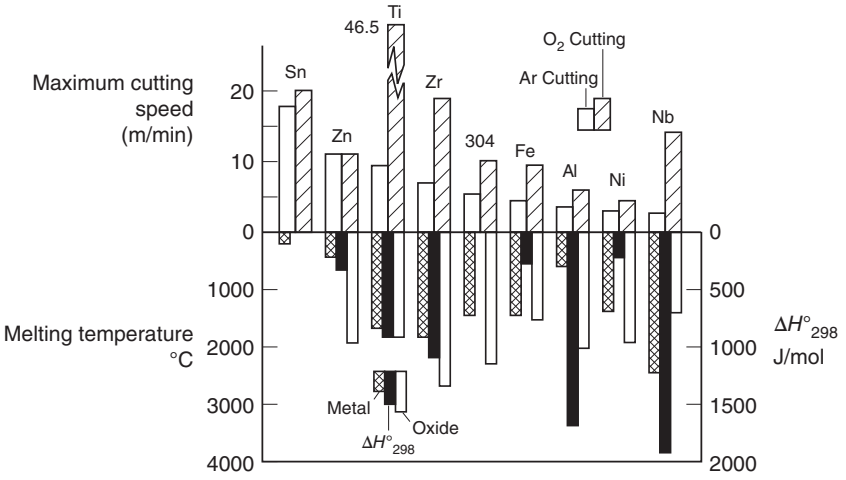


Figure 9.7 Comparison of maximum cutting speed with O_2 and Ar assist gases for different metals. $\Delta H°$ is the energy associated with exothermic reaction of the process. Source: From Miyamoto and Maruo (1991)/with permission of Elsevier.

cutting of titanium, zirconium, and niobium due to the relatively high exothermic energy associated with these metals. However, the surface quality obtained is relatively poor. This is due to the fact that the oxidation region cannot be limited to the beam irradiating region as a result of the high exothermic energy. The relatively low cutting speeds achieved with aluminum and zinc is due to the high melting temperature of their oxides. The exothermic reaction associated with steel is not very high, when compared to that of other metals such as titanium, zirconium, and niobium. Thus, the cutting speeds are not very high, but the quality of cut is relatively good since the relatively low exothermic reaction results in a reaction zone that is limited to the beam-irradiated region.

Care must be used in oxygen-assisted cutting, since excess oxygen may result in overreaction or uncontrollable burning away from the main cutting direction, especially for thick materials. That may increase striation formation, and thus, roughness.

Gas Nozzles Some of the nozzle designs that are commonly used for coaxial application of a gas jet during laser cutting are shown in Figure 9.8.

Low or subsonic flow rates from a coaxial nozzle are found to produce repeatable results, especially when the nozzle is positioned close to the workpiece, that is, with a stand-off (nozzle to work) distance of about 0.1–1.5 mm. Short stand-off distances are used despite the fact that the effective cutting pressure decreases rather slowly with distance from the nozzle tip (about 15% decrease over a nozzle gap of 10 mm). This is because the jet direction is highly sensitive to manufacturing

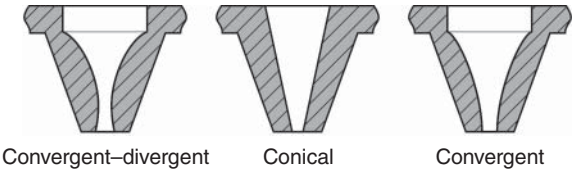


Figure 9.8 Nozzle designs for laser cutting. Source: From Fieret et al. (1986)/SPIE.

imperfection or tip damage. Even though higher pressures or flow rates result in higher cutting speeds and quality, they do not necessarily ensure repeatability of the process due to turbulence in the gas flow.

9.1.3.5 Effect of Focal Position Due to divergence of the beam, it is essential to have consistency in the location of the focal point relative to the workpiece surface. Best results (minimum kerf width) are obtained when the focal point is either on the workpiece surface or just below it (Figure 9.9). For thick plates, it might be preferable to have the focus positioned one-third of the plate thickness below its surface.

9.1.4 Laser Cutting Principles

The laser fusion cutting process is illustrated schematically in Figure 9.1. Both the laser beam and gas jet impinge on the workpiece surface. Sources of heat for the cutting process include

1. Absorbed laser radiation.
2. Energy due to exothermic reaction between the base material and assist oxygen gas.

The energy produced melts and may partly evaporate material in front of the beam. The pressurized gas jet then ejects the molten material from the lower surface of the workpiece. Material removal thus occurs by

1. Evaporation from the surface of the molten layer.
2. Ejection from the lower surface of the workpiece due to friction between the gas flow and the surface of the molten layer.

Energy is lost from the process by

1. Heat conduction.

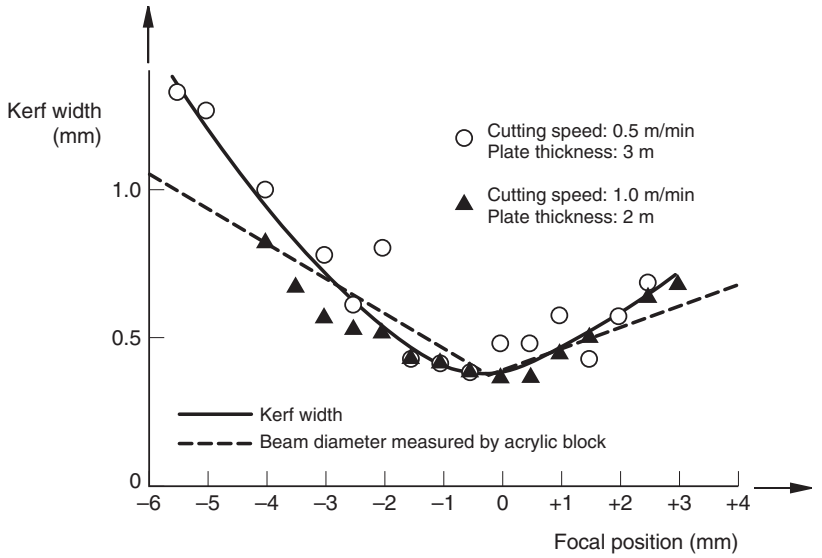


Figure 9.9 Effect of Focal position on kerf width for a 7075-T6 aluminum alloy cut with a 1.8 kW CO_2 laser power and argon assist gas. Source: From Masumoto et al. 1992/Japan Welding Society.

2. Evaporation from the erosion front.
3. Melting of solid metal.
4. Ejection of the molten metal.
5. Reflection, radiation, and convection cooling by the gas flow. For subsonic gas flow, the convection cooling effect is considered negligible.

In the initial stages of the process, the entire beam and gas jet hit the workpiece surface. However, once the process is initiated, a kerf is formed, and only a portion of the gas and laser beam impinge on the top surface of the workpiece directly ahead of the kerf, and that portion may be reflected back. The remaining portion of the beam propagates downward into the kerf and is partly absorbed at the front end of the kerf (erosion front), which is slightly inclined to the vertical. For a beam that is linearly polarized in the cutting direction, most of the beam power is absorbed at the erosion front. A thin layer of molten material forms at the interface between the kerf and the solid base material in front of it and is ejected by the mechanisms listed earlier. The thickness of the melt increases with increasing cutting speed, since more molten metal is then produced per unit time. However, an increase in assist gas velocity decreases the melt thickness, since it results in a more rapid ejection of molten material (Figure 9.10).

For a given laser power, the average temperatures attained in the molten layer decrease as the thickness of the workpiece increases, until above a certain thickness, the average temperature falls below the melting temperature. This is the maximum thickness that can be cut with a laser of that power level, all other laser parameters being the same.

9.1.4.1 Beam Absorption During Laser Cutting Absorption of the laser beam during cutting may be enhanced by a number of phenomena, including surface roughness, oxide formation, and plasma formation. Absorption may be by Fresnel absorption or by inverse bremsstrahlung (Section 10.3.2.1). However, since the direction of beam polarization strongly influences the cutting efficiency, this suggests that Fresnel absorption is the principal mechanism. Since a significant portion of the energy generated is removed with the ejected molten metal, temperatures in the cutting region may not be

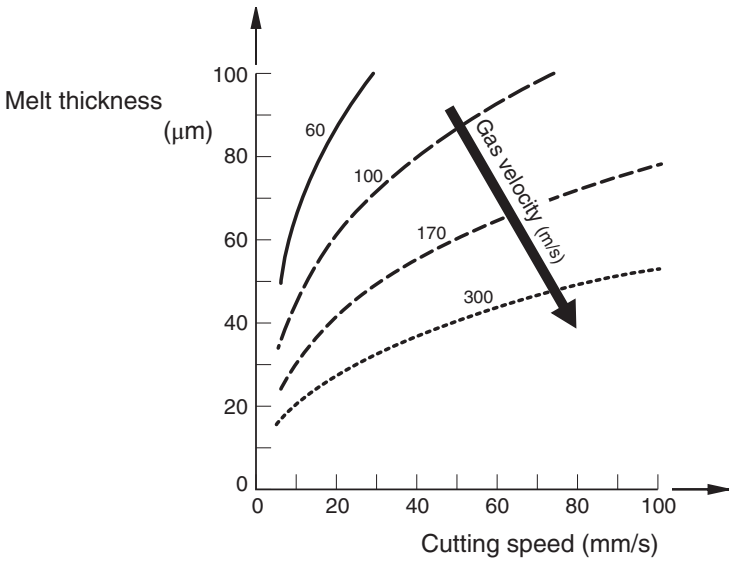


Figure 9.10 Schematic variation of melt thickness with cutting speed. Source: From Vicanek et al. (1987)/IOP Publishing.

high enough to generate the high vapor densities essential for plasma formation. The role of the plasma in the cutting process (and therefore absorption by inverse bremsstrahlung) is thus minimal, as opposed to the case of welding and drilling.

For p-polarized light (plane-polarized light parallel to the plane of incidence) which is incident on metallic surfaces, the maximum value of the absorption coefficient occurs at high angles of incidence, θ (angle between the laser beam and the surface normal). The laser beam is then almost parallel to the surface on which it is incident (Figure 9.11). This is supported by the fact that the inclination of the cut surface on which the laser beam is incident is such that the angle of incidence is only a few degrees below 90°. The absorption is also about an order of magnitude greater than that at normal incidence (when the laser beam is normal to the surface on which it is incident). Furthermore, a p-polarized beam is absorbed much more strongly than an s-polarized light. Thus, a p-polarized beam is more efficient for laser cutting in a straight line.

Absorption efficiency of the laser beam in cutting varies with the beam intensity, being low at low intensities, increasing with increasing beam intensity to a maximum, and then decreasing with further increases in the intensity (Figure 9.12). This is due to the dependence of absorption on the angle of incidence. At low intensities, the cutting front is relatively flat, and that results in low absorption since the angle of incidence is then almost zero, that is, the laser beam is almost normal to the workpiece surface. With increasing intensity, the angle of incidence increases, further increasing the absorption. At very high intensities, the cutting front is almost vertical, resulting in almost 90° incidence, at which the absorption is relatively low (Figure 9.13). For typical cutting conditions, the inclination of the cutting front to the vertical is less than 1° close to the top of the workpiece and increases to an inclination of 3–5° toward the bottom of the workpiece.

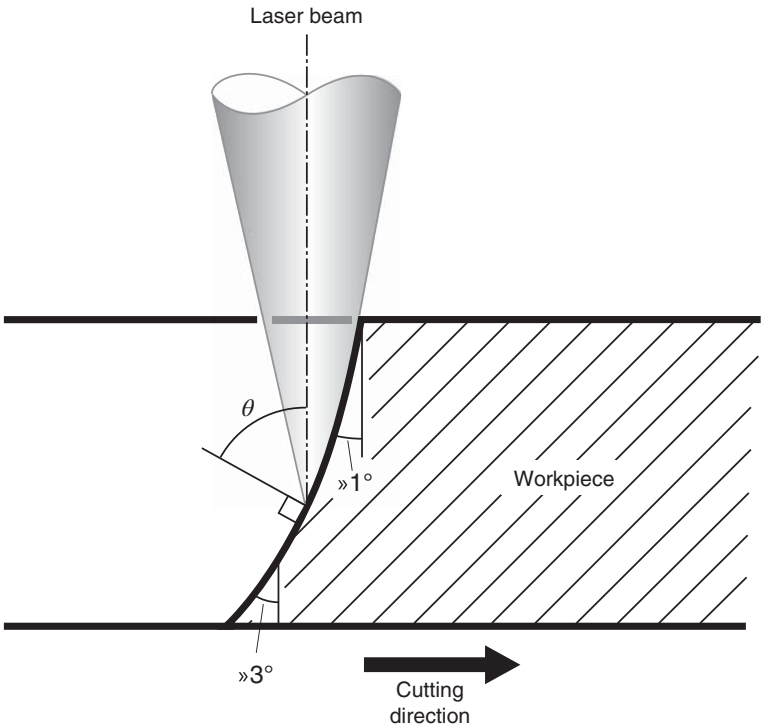


Figure 9.11 Angle of incidence of laser beam on cut surface, and inclination of the cut surface with the vertical.

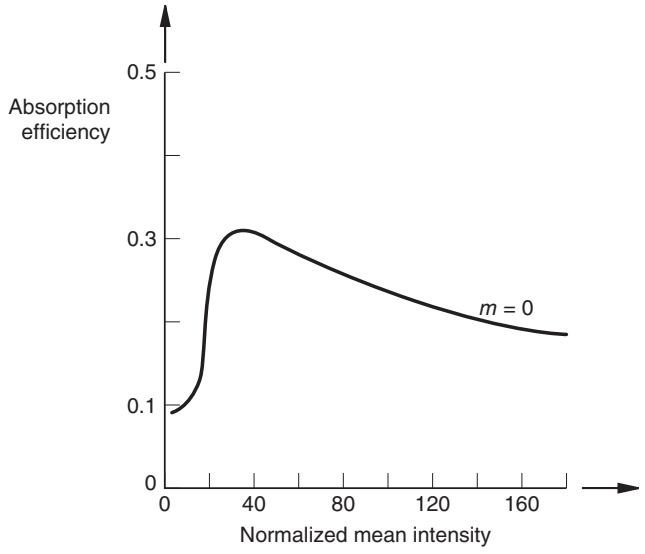


Figure 9.12 Variation of absorption efficiency of a p-polarised beam with beam intensity for a Gaussian beam ($m = 0$). Source: From Schulz et al. (1987)/IOP Publishing.

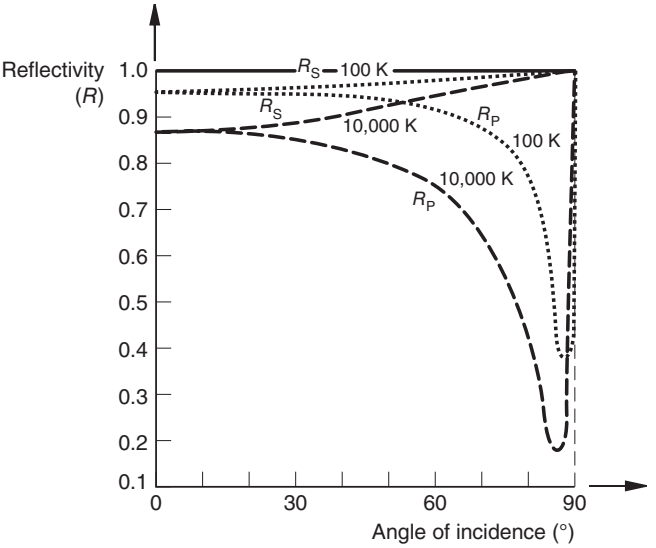


Figure 9.13 Variation of reflectivity of mild steel with beam incidence angle. R_p is the reflectivity of p-polarized light, and R_s is the reflectivity of s-polarized light. Source: From *ICALEO 1991 Proceedings*, Vol. 74. Copyright 1992, Laser Institute of America. All rights reserved.

9.1.4.2 Process Modeling The laser cutting process is modeled in this section to relate process inputs such as gas velocity, cutting velocity, laser power, and material properties to the process outputs such as the melt thickness and quality of the cut surface. The gas jet in laser cutting significantly influences the process dynamics. Modeling of the process thus involves consideration of the following:

1. Gas flow.
2. Flow of the molten metal.
3. Heat flow in the solid material.

These flow regimes are now addressed in the following sections.

Gas Flow Model To provide a basic understanding of the process without extensive analysis, we consider a nonreactive inert assist gas, thereby neglecting exothermic reaction, and make the following simplifying assumptions:

1. Gas flow is subsonic and laminar.
2. The flow regime consists of a region of irrotational motion and a boundary layer region at the cutting front.
3. The cutting front is an inclined plane.
4. The flow only separates at the lower edge.
5. The gas stream is bounded by free streamlines. This is an oversimplification since the gas is surrounded to a large extent by air of almost the same density as the gas. Turbulent mixing of the gas stream with the surrounding air increases downstream from the upper edge. Thus for relatively thick workpieces, the assumption will only hold for the upper portion of the cutting front.

With this background, and considering the schematic in Figure 9.14, we can obtain an expression for the pressure distribution, P_g, in the gas stream using Bernoulli's equation:

$$\frac{P_1}{\rho g_1} + \frac{{u_1}^2}{2g} + z_1 = \frac{P_2}{\rho g_1} + \frac{{u_2}^2}{2g} + z_2 + \text{friction} \tag{9.1a}$$

where g is the acceleration due to gravity (m/s^2), P is the pressure at a given location (Pa), u is the velocity at that location (m/s), z is the height of the location of interest (m), and subscripts 1 and 2 refer to two different locations.

Neglecting friction and the small height difference in the laser cutting area and considering the reference gage pressure at the nozzle to be zero (atmospheric), we have

$$P_g = \frac{1}{2}\rho_g {u_{g0}}^2 - \frac{1}{2}\rho_g {u_g}^2 \tag{9.1b}$$

where P_g is the pressure distribution in the gas stream, u_g is the gas velocity distribution, u_{g0} is the gas velocity at the nozzle, ρ_g is the room temperature gas density, which is assumed to be constant with pressure (i.e. incompressible).

The velocity distribution is needed to obtain the pressure distribution from Eq. (9.1b). For the shear stress distribution, we first consider the coordinate system shown in Figure 9.14 and then make the following assumptions:

1. Flow is Newtonian.
2. The thickness of the molten layer can be neglected.

Now considering the fact that the shear stress is proportional to the shear strain rate since flow is Newtonian, the shear stress distribution at the cutting front can be expressed as:

$$\tau = \mu_g \left.\frac{\partial u_{gz}}{\partial x}\right|_{x=0} \tag{9.2}$$

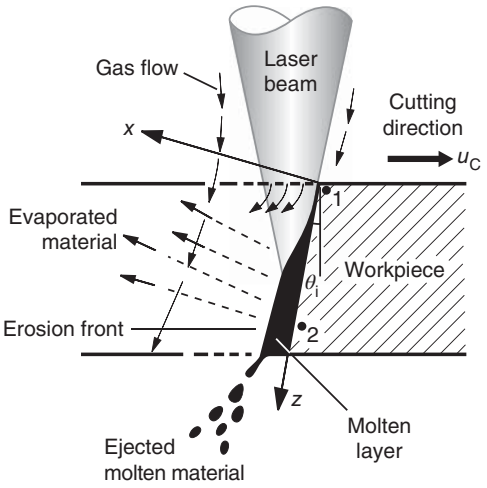


Figure 9.14 A two-dimensional view of the cutting process. Source: From Vicanek and Simon (1987)/IOP Publishing.

The velocity field is obtained from the boundary layer equation:

$$\rho_g \left(u_{gz} \frac{\partial u_{gz}}{\partial z} + u_{gx} \frac{\partial u_{gz}}{\partial x} \right) = -\frac{\partial P_g}{\partial z} + \frac{\partial}{\partial x} \left(\mu_g \frac{\partial u_{gz}}{\partial x} \right) \tag{9.3}$$

where μ_g is the gas viscosity (Pa s), τ is the shear stress distribution at the cutting front (N/m^2), u_{gz} is the gas velocity in the z-direction, and u_{gx} is the gas velocity in the x-direction.

The pressure gradient in Eq. (9.3) is assumed to be known. The resulting pressure and shear stress distributions along the cutting front obtained from Eqs. (9.1b)–(9.3) are of the form illustrated in Figures 9.15 and 9.16, respectively. The pressure distribution has a peak just below the upper edge of the cutting front and then reduces monotonically to zero at the lower edge. The pressure also increases with increasing inclination (θ_i) of the cutting front from the vertical. The smaller the cutting front slope, the lower the resistance to flow, and thus the smaller the pressure. The shear stress, on the other hand, remains relatively constant over most of the cutting front and is less sensitive to the inclination.

Modeling the Melt Flow Since most of the molten material is forcibly ejected by the gas jet, only a thin layer of melt remains on the cutting front, with a thickness of the order of 40 μm. Thus flow may be assumed, with a reasonable degree of accuracy, to be two-dimensional. In other words, this is a boundary layer problem. Using the coordinate system shown in Figure 9.14 and assuming incompressible flow, the continuity equation can be expressed as:

$$\frac{\partial u_{lz}}{\partial z} + \frac{\partial u_{lx}}{\partial x} = 0 \tag{9.4}$$

Likewise, the Navier–Stokes equation is given by:

$$\rho_l \left(\frac{\partial u_{lz}}{\partial t} + u_{lx} \frac{\partial u_{lz}}{\partial x} + u_{lz} \frac{\partial u_{lz}}{\partial z} \right) = -\frac{\partial P_l}{\partial z} + \mu_l \left(\frac{\partial^2 u_{lz}}{\partial x^2} + \frac{\partial^2 u_{lz}}{\partial z^2} \right) \tag{9.5}$$

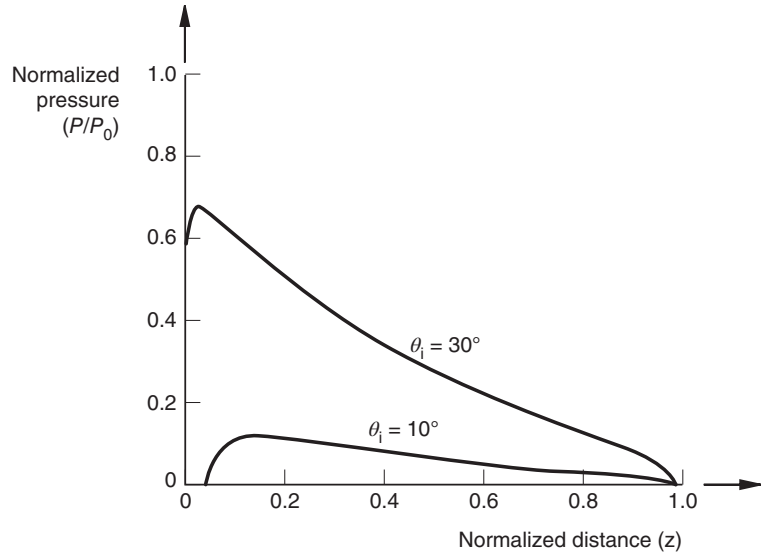


Figure 9.15 Schematic of pressure distribution along the cutting front. P_0 is the reference pressure at the stagnation point. Source: From Vicanek and Simon (1987)/IOP Publishing.

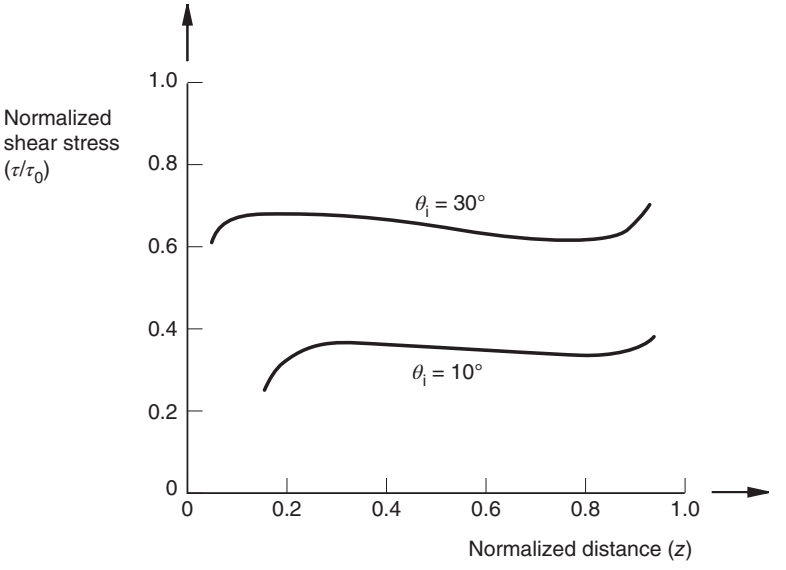


Figure 9.16 Schematic of shear stress distribution along the cutting front. τ_0 is a reference shear stress. Source: From Vicanek and Simon (1987)/IOP Publishing.

where P_l is the pressure in the molten metal and is assumed to be independent of x since this is a boundary layer problem, u_{lz} and u_{lx} are the molten material velocities in the z and x directions, respectively, μ_l is the dynamic viscosity of the molten metal, and ρ_l is the density.

Boundary conditions have to be specified at the solid–liquid interface, $x = 0$, and also at the surface of the liquid layer, $x = h_x(z, t)$, where h_x, the melt thickness, is a function of both position and time.

At the solid–liquid interface, material is uniformly moved with the cutting velocity u_c (Figure 9.14), and thus continuity there requires that

$$u_{lz} = u_c \sin\theta_i \qquad \text{at} \qquad x = 0 \tag{9.6a}$$

$$u_{lx} = u_c \cos\theta_i \qquad \text{at} \qquad x = 0 \tag{9.6b}$$

where u_c is the cutting velocity and θ_i is the inclination of the cut surface with the vertical.

There are normal and tangential stresses, P_g and τ, respectively, at the melt surface due to the gas jet. For equilibrium, we have

$$P_l = P_g + \frac{\gamma_l}{r_c} \qquad \text{at} \qquad x = h_x \tag{9.7a}$$

$$\mu_l \frac{\partial u_{lz}}{\partial x} = \tau_{gl} \qquad \text{at} \qquad x = h_x \tag{9.7b}$$

where h_x is the melt thickness, r_c is the local radius of surface curvature, γ_l is the surface tension coefficient of the molten material, and τ_{gl} is the shear stress due to friction at the interface between the gas and molten material.

The following kinematic condition ensures that there is no mass transfer through the surface:

$$\frac{\partial h_x}{\partial t} + u_{lz}\frac{\partial h_x}{\partial z} = u_{lx} \qquad \text{at} \qquad x = h_x \tag{9.8}$$

Evaporation effects are neglected in Eq. (9.8) for simplicity. Other conditions that can be applied are as follows:

$$u_{lz} = 0 \qquad \text{when} \qquad t = 0 \tag{9.9a}$$

$$h_x = 0 \qquad \text{when} \qquad t = 0 \tag{9.9b}$$

$$h_x = 0 \qquad \text{at} \qquad z = 0 \tag{9.9c}$$

Knowing P_g and τ from the solution to the gas flow equations, Eqs. (9.4)–(9.9) can be solved numerically for the pressure, velocity, and height of the molten layer. Typical trends obtained from solution of these equations are illustrated in Figure 9.10, which shows that increasing the cutting velocity u_c increases the melt thickness h_x, and that increasing the gas jet velocity decreases the melt thickness.

The Heat Flow Effect The analyses in the preceding sections neglect the effect of heat flow. A more complete solution of the pressure and stress fields requires consideration of the thermal effect. The governing equation for heat flow is given by Eq. (4.2) as:

$$\rho c_p \frac{dT}{dt} = \frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial y}\left(k\frac{\partial T}{\partial y}\right) + \frac{\partial}{\partial z}\left(k\frac{\partial T}{\partial z}\right) + q_s \tag{4.2a}$$

In Chapter 4, this equation was first simplified by neglecting the internal energy generation term, q_s. The resulting equations associated with the line heat source model, Eqs. (4.22), (4.24), and (4.31), for the temperature distribution, peak temperatures, and cooling rates, are still applicable as initial approximations for conduction heat flow in the workpiece when an inert gas or nitrogen is used as the assist gas. When oxygen or air is used as the assist gas, however, q_s cannot be neglected, as the exothermic reaction is then significant enough to influence the resulting temperatures.

For a more accurate analysis, however, the continuity equation, Navier–Stokes equation, and the thermal equation will have to be solved numerically, with the appropriate boundary conditions for the gas flow, melt flow, and solid workpiece regimes.

Example 9.1 A 2.5 mm thick aluminum plate is cut using a 1.5 kW laser. If the resulting kerf width is 0.3 mm; the ambient temperature is 25 °C; and 5% of the incident laser beam is absorbed by the plate, determine the cutting speed. You may assume that there is no vaporization, and that there are no energy losses by conduction, convection, or radiation.

Solution:
We apply the discussion in Chapter 4 to this problem. The material properties are obtained from Appendices 4.D and 4.E as:

Average density, $\rho = 2700\,\text{kg/m}^3$
Average specific heat, $c = 900\,\text{J/kg K}$
Latent heat of fusion, $L = 397\,\text{kJ/kg}$
Melting temperature, $T_\text{m} = 660.4\,^\circ\text{C}$

From Eq. (4.1), we have

$$Q_\text{a} = q \times t = m_\text{a} c_\text{p} \Delta T_\text{m} + m_\text{a} L_\text{m} + m_\text{a} c_\text{p} \Delta T_\text{v} + m_\text{a} L_\text{v} + Q_\text{l}$$

where q is the power supplied and t is the cutting time. Since it is assumed that there is no vaporization, and that there are no energy losses by conduction, convection, or radiation, we have

$$q \times t = m_\text{a} c_\text{p} \Delta T_\text{m} + m_\text{a} L_\text{m} = \rho V (c_\text{p} \Delta T_\text{m} + L_\text{m})$$

or

$$q \times t = \rho \times l \times w_\text{k} \times h (c_\text{p} \Delta T_\text{m} + L_\text{m})$$

where l is the length of the cut that is made, w_k is the kerf width, and h is the plate thickness. Thus, the average cutting speed, u_c, is

$$\begin{aligned} u_\text{c} = \frac{l}{t} &= \frac{q}{\rho \times w_\text{k} \times h (c_\text{p} \Delta T_\text{m} + L_\text{m})} \\ &= \frac{1500 \times 0.05}{2.7 \times 10^{-6} \times 0.3 \times 2.5 (900 \times (660.4 - 25) + 397 \times 10^3)} \\ &= 38.23\,\text{mm/s} = 2.3\,\text{m/min} \end{aligned}$$

9.1.5 Quality of Cut Part

The major factors that determine the quality of the cut part are the surface striations, dross formation, and cracking. In the following sections, we discuss the mechanisms of striation and dross formation. Cracking is discussed in relation to specific materials in Section 9.1.6 and also more extensively in Section 5.2.2.

9.1.5.1 Striations of the Cut Surface The surfaces cut using a laser beam normally have a nearly periodic striation pattern (Figure 9.17a), that results in a surface roughness which may also vary in the thickness direction (Figure 9.17b), depending on the processing conditions.

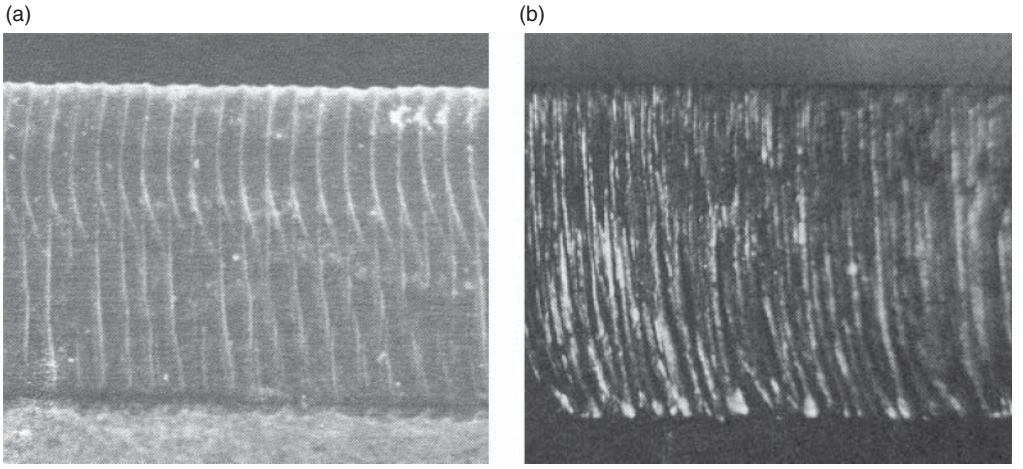


Figure 9.17 Illustration of striations formed on a surface after laser cutting. (a) Periodic nature of the striation pattern. (b) Variation of striation pattern in the thickness direction for relatively thick materials. Source: From (a) Ready (2001), with permission from Springer Nature; (b) Powell et al. (1988)/SPIE.

Primary causes of these striations are the following:

1. Vibrations in the motion unit.
2. Fluctuations in the laser power.
3. Fluctuations in the gas flow.
4. Hydrodynamics of the molten metal flow.

A unique aspect of the striations is the two distinct patterns evident on the surface (Figure 9.15b). One pattern is closer to the upper surface of the workpiece, with relatively fine striations. The other pattern is on the lower portion of the cut surface, with relatively coarse striations. The two patterns are separated by a distinct line that is almost parallel to the workpiece surface. The two striation patterns result from the temperature distribution in the molten layer in the vertical direction, being higher in the upper portion compared to the lower portion.

The speed of the assist gas as it enters the kerf is generally subsonic. However, its temperature increases as it penetrates deeper into the kerf, resulting in a decrease in the gas density. In a sense, then, the cut acts as a converging nozzle, increasing the gas flow velocity. When the flow reaches sonic velocities, it becomes turbulent and thus changes the dynamics of the molten layer. The turbulence in the gas flow also cools the molten layer more effectively, and since the oscillation frequency of the molten layer when subjected to perturbations is higher for higher temperatures, the frequency of the striation pattern is lower for the lower portion of the cut surface. The characteristic frequencies of the oscillations depend on the workpiece thickness and may be of the order of 10^3 cycles/s, increasing with cutting speed. As the thickness of the cut piece increases, the striations tend to become irregular, and the average roughness of the cut surface increases (Table 9.3).

The increase in roughness of the lower portions of the cut surface with workpiece thickness is also partly due to the inability to provide adequate energy on the lower portion of the workpiece to produce molten metal of very low viscosity.

9.1.5.2 Dross Formation Dross is essentially material that clings to the lower edge of the workpiece and appears as solidified drops after laser cutting (Figure 9.18). The formation of dross

Table 9.3 Illustration of roughness variation of cut surface with workpiece thickness for a steel plate.

Thickness (mm)	Upper roughness (μm)	Lower roughness (μm)
1	3	3
3	9	11
5	24	30
7	40	55
9	70	60
10	85	110

Source: Table 3, Schoucker (1986)/SPIE.

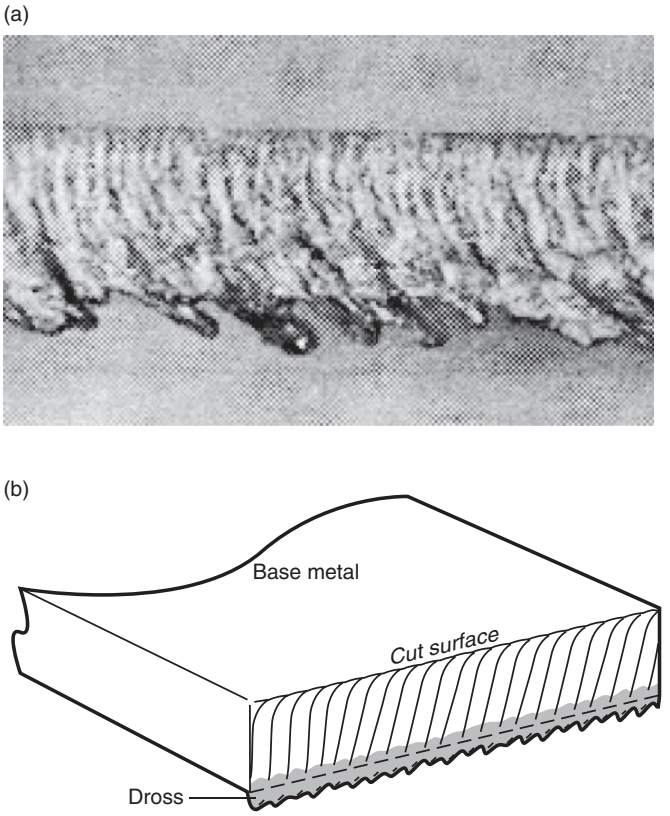


Figure 9.18 Dross formed on the lower edge of the workpiece after laser cutting. (a) Photograph of dross. (b) Schematic of dross. Source: (a) From Powell et al. (1988)/SPIE.

depends on the surface tension and viscosity of the molten material. The higher the surface tension or viscosity, the greater the tendency to form dross, since that prevents smooth flowing of the molten material out of the reaction zone. Inert gas-assisted cutting has a greater tendency to form dross as compared to oxygen-assisted cutting of the same material, since the surface tension of the pure metal is generally greater than that of its oxide. Thus, higher pressures (over 1 MPa or 10 bars) are required to achieve dross-free cutting with inert gas assist than with oxygen assist. Furthermore,

the dross formed during inert gas-assisted cutting is usually more difficult to remove than the more brittle oxide dross formed when oxygen assist is used.

9.1.6 Material Considerations

The effectiveness of a laser in processing a given material depends on the properties of the material, for example absorptivity (1 – reflectance) at that wavelength; thermal conductivity; melting and boiling points; heat of reaction; and surface tension of the molten material. The reflectance, thermal, and physical properties of some common metals are summarized in Appendices 4.D–4.F. In Sections 9.1.6.1–9.1.6.2, we consider the behavior of both metals and nonmetals when cut with a laser.

9.1.6.1 Metals

Plain Carbon Steels Oxygen-assisted cutting is normally used for carbon steels where the exothermic reaction between oxygen and iron aids in the cutting process, resulting in cutting speeds which are much higher than those achieved when inert gas is used. The exothermic reaction is given by the following equation:

$$\mathrm{Fe} + \frac{1}{2}\mathrm{O_2} = \mathrm{FeO} + \text{Energy}\ \ (258\ \mathrm{kJ/mol}) \quad \text{at} \quad 2000\ \mathrm{K} \tag{9.10}$$

$$2\mathrm{Fe} + \frac{3}{2}\mathrm{O_2} = \mathrm{Fe_2O_3} + \text{Energy}\ \ (827\ \mathrm{kJ/mol}) \quad \text{at} \quad 2000\ \mathrm{K} \tag{9.11}$$

Oxygen-assisted cutting also helps reduce dross formation, since the viscosity of the molten metal decreases when oxidized, making it easier to be removed by the gas jet. The presence of phosphorus and sulfur in the carbon steel can cause burnout along the cut edge. An increase in carbon content of the steel tends to improve the edge quality, but with a greater tendency for cracking in the HAZ (Section 5.2.2).

Galvanized Steel Laser cutting of galvanized steel often results in rough edges with extensive dross. This is most likely due to the formation of zinc oxide which does not flow easily. To obtain reasonable quality, lower cutting speeds are normally used, compared to the speeds used for uncoated steels.

Stainless Steels The small HAZ associated with laser cutting minimizes any impact that the cutting process may have on its corrosion resistance. In using oxygen assist, the heat of reaction that results is not as significant as that associated with plain carbon steels. Thus, the cutting speeds achieved for stainless steels are relatively lower than those for carbon steels. As a result, oxygen assist is normally not recommended for cutting of stainless steels, especially since the burning caused by the assist gas may be detrimental to the cut quality. Ferritic and martensitic stainless steels tend to produce clean and smooth-edged cuts, while austenitic stainless steels tend to have dross sticking out from the lower surface of the workpiece. This is primarily due to the high viscosity of molten nickel.

Alloy and Tool Steels The edges produced in laser cutting of high alloy steels such as AISI 4340 steel are often clean and square. This is essentially due to the more precise control of the alloying element contents of such materials. Likewise, a number of tool steels are also good candidates for laser cutting. However, the tungsten-based tool steels, Groups T and H, retain a lot of heat in the molten state, resulting in often burned out cuts.

Aluminum and Its Alloys Aluminum and its alloys are highly reflective (up to 97%) of the high 10.6 μm CO_2 laser beam wavelength. This, coupled with their relatively high thermal conductivity (≈247 W/m K), makes it difficult to initiate cutting in aluminum alloys. High beam intensities,

achieved using either superpulsing or hyperpulsing (Section 8.1.4), are thus necessary to cut aluminum alloys. Absorption of the laser beam at the top surface can be enhanced by using any of the coating techniques discussed in Section 11.1.2.3, but anodizing aluminum has been found to be effective. This is done by coating the base material with a thin layer (approximately 20 μm) of Al_2O_3. Al_2O_3 is highly absorptive (≈100%) of the infrared CO_2 laser beam at thicknesses greater than 5 μm. Absorptivity can also be improved by graphite coating.

Aluminum undergoes a highly exothermic reaction with oxygen as:

$$4Al + 3O_2 = 2Al_2O_3 + \text{Energy}\ \ (1670\,\text{kJ/mol}) \quad \text{at} \quad 293\,\text{K} \tag{9.12}$$

which can increase cutting speeds to some extent. However, the exothermic reaction is not as effective in increasing the cutting speeds for the same amount of released energy as compared to the case of iron. This is because the resulting oxide layer, aluminum oxide (Al_2O_3), reduces the diffusion of atoms to the reaction front, thereby reducing the oxidation rate. The oxide layer is periodically disrupted during the cutting process by the pressurized oxygen jet, permitting some amount of oxidation to occur. The function of assist gas in cutting these materials is thus primarily to eject the molten metal. As a result, oxygen assist is normally not recommended for cutting aluminum.

Dross formation can be a problem in cutting aluminum. This is because the high thermal conductivity of the base material results in a low temperature melt. This results in the surface tension of the molten layer being relatively high, and thus more difficult to eject, leaving dross at the bottom edge of the cut. The dross formed is easily removed by mechanical means, though. The high thermal conductivity of the base material may sometimes lead to overheating of small components during cutting. This, combined with the low melting temperature may even result in widescale melting. This can be prevented by spraying either water or soluble oil on the component during cutting. The water is kept away by the pressure of the gas jet and cools the component by evaporation.

Laser cutting sometimes produces intergranular cracks (cracks that preferentially propagate along grain boundaries) on the surface of some aluminum alloys, making them unsuitable for aircraft structural components.

Copper and Its Alloys Both the conductivity (≈398 W/m K) and reflectivity (up to 98.4%) of copper are higher than those of aluminum, making it even more difficult to process with a CO_2 laser. Even though the exothermic reaction associated with the process can increase the cutting speeds, the energy released is relatively small:

$$2Cu + \frac{1}{2}O_2 = Cu_2O + \text{Energy}\ (167\,\text{kJ/mol}) \quad \text{at} \quad 293\,\text{K} \tag{9.13}$$

$$Cu + \frac{1}{2}O_2 = CuO + \text{Energy}\ (155\,\text{kJ/mol}) \quad \text{at} \quad 293\,\text{K} \tag{9.14}$$

Any increase in cutting speed is primarily due to increased absorption that results from the formation of the highly absorptive oxide layer in the vicinity of the cut.

A number of copper alloys such as brass and bronze are more effectively processed due to a reduced reflectivity and thermal conduction.

Titanium and Its Alloys Titanium is highly effective in absorbing the CO_2 laser beam. The exothermic reaction between oxygen and titanium when oxygen assist gas is used greatly enhances the cutting efficiency, thereby increasing the cutting speeds. The reaction with oxygen may be expressed as:

$$Ti + O_2 = TiO_2 + \text{Energy}\ (912\,\text{kJ/mol}) \quad \text{at} \quad 293\,\text{K} \tag{9.15}$$

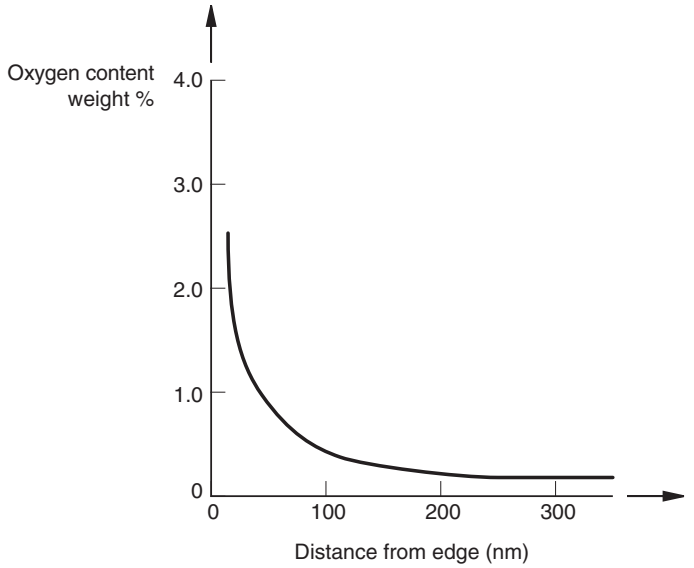


Figure 9.19 Illustration of oxygen content variation with distance from the cut edge in laser cutting of a titanium alloy, Ti_6Al_4V, using high-pressure argon assist gas. Source: From Powell et al. (1988)/SPIE.

A relatively thick and brittle oxide layer (of the order of 1–3 mm) is normally produced along the cut edge, since the reaction is very rapid at the high cutting temperatures. When a Ti_6Al_4V titanium alloy is cut with a laser using high-pressure inert gas, the resulting variation of oxygen content with depth from the cutting edge is illustrated in Figure 9.19.

This process is mainly useful in applications where the oxide layer is not detrimental, such as low service life or electrical applications where titanium may be used for corrosion resistance, rigidity, and its lightweight characteristics. However, for aerospace applications where they are extensively used for their toughness, fatigue resistance, and high strength-to-weight ratio, these properties are highly degraded by the oxide layer formed. Cutting titanium with oxygen assist for such applications is therefore undesirable. Thus, it may be preferable to use an inert gas assist when cutting titanium alloys for aerospace applications, but at a reduced cutting rate which may be an order of magnitude lower than that attainable when using oxygen-assist. On the other hand, there is a tendency for dross to form on the underside of the cut edge under such circumstances. As an example, high cutting speeds of about 3.6 m/min can be obtained in cutting a 6.4 mm thick plate with 500 W power using oxygen-assist.

Nickel and Its Alloys Pure nickel has a relatively low thermal conductivity (≈59 W/m K) and reflectivity (up to 94%) compared to those of copper and aluminum, making it easier to cut using lasers. The use of oxygen assist does improve cutting speeds as a result of the exothermic reaction:

$$\mathrm{Ni} + \frac{1}{2}\mathrm{O_2} = \mathrm{NiO} + \text{Energy (244 kJ/mol)} \quad \text{at} \quad 293\ \mathrm{K} \tag{9.16}$$

This is more effective when cutting Ni–Fe alloys. Some nickel alloys, however, have reflectivities and conductivities that make them more difficult to process with lasers. Furthermore, the high viscosity of molten nickel results in dross formation.

Despite the relative ease with which a number of nickel alloys can be processed with a laser, the metallurgical effects of the process, even in the small HAZ, precludes its use in some aerospace applications such as jet engines where the working environment can be aggressive.

9.1.6.2 Nonmetals

Polymers In cutting polymers, the intense laser beam tends to break down the polymer chains. Edges produced in cutting thermoplastics appear polished as a result of resolidified melting. As the strength of the polymer increases, there is a tendency for charring to occur along the cut edge, since more energy is then required to break down the bonds. Polyester and polycarbonate are relatively easier to cut, while with PVC (polyvinyl chloride), phenolic, and polyimide, a significant amount of decomposed material may be found along the cut edge. Care must always be taken to contain hazardous and/or corrosive fumes that may be generated in the processing of polymers. It is generally preferable to use air rather than oxygen in cutting polymeric materials.

Both natural and synthetic rubber are effectively cut in thicknesses up to 19 mm using lasers, and they tend to exhibit slight stickiness along the edge when freshly cut.

Composites Since composites consist of different materials, the cutting conditions are determined by only one of the component materials, which may result in degradation of the other materials during cutting. One approach to this problem is to cut at higher speeds using higher-power lasers. Thin sheets (say, less than 0.5 mm) of laminated composites with a polymeric material as the matrix can be trimmed using lasers before they are cured. Thicker sections or fully cured composites often result in charring, delamination, and thermal damage along the cut edge. When the properties of the constituent materials are similar, the composite material is then easier to process using a laser. In all cases, care must also be exercised in containing hazardous fumes and material.

Carbon fiber-reinforced plastic (CFRP) composites are being increasingly used in the aerospace (aircraft wings) and automotive (panels) industries. Cutting the material is a necessary step in production of components. However, with conventional machining operations, the abrasive nature of the carbon fibers results in rapid tool wear and thus high processing costs. In addition, high contact forces may cause delamination during machining and abrasive waterjet cutting. With lasers being noncontact, they have an advantage in that respect. On the other hand, an issue of concern in using lasers is the resulting HAZ (Section 5.1.3). Other defects that can result with lasers include microcracks, delamination, cavities, and protruding fibers. The HAZ size is found to be greater in the direction parallel to the fibers because of the relatively high fiber conductivity. Since the energy per unit length is smaller at higher speeds, multi-pass operations are often used (especially for thick materials) at high speeds to reduce the overall HAZ size.

Quartz and Glass The relatively low coefficient of expansion of quartz makes it easier to cut in comparison with glass. The high coefficient of thermal expansion of glass results in thermal shock that tends to form edge cracks.

One approach to cutting such brittle materials is by first laser scribing to remove a small amount of material from the surface of the glass, and then applying force by flexing, to fracture the material along the line of scribing. This requires only a small power density, resulting in a very small HAZ size. Another approach involves controlled thermal fracture, which takes advantage of the localized thermal stresses that are set up due to the high coefficient of thermal expansion and high beam intensity to fracture the material. The process is quite controllable, with the fracture essentially following the path of the beam.

Ceramics The high transverse temperature gradients associated with laser processing can generate relatively high thermal stresses that cause cracking in ceramic materials, which are inherently brittle. Furthermore, the thermal conductivity decreases rapidly with increasing temperature in such materials, significantly reducing thermal diffusion into the body of the workpiece. Preheating the workpiece to a relatively high temperature before cutting or preheating the region directly ahead of the laser beam reduces the transverse temperature gradient, and thus the tendency to crack.

Cutting of alumina often results in cracking. Cracks also develop during cutting of silicon carbide of thickness greater than 3 mm. Silicon nitride, on the other hand, can be cut to 4 mm thickness at about 10 mm/s. The surface roughness obtained when zirconia is cut is about 10 μm, while silicon nitride results in a surface roughness of about 100 μm.

Textiles and Fabrics Lasers are also used in the textile industry to cut fabric. To increase productivity, several layers of fabric may be stacked up and cut at the same time. However, this causes the edges of the individual layers to become welded together. This may be prevented by pressing the layers together to prevent the cutting gas from entering sideways between them.

Wood Wood can generally be cut using CO_2 lasers. The cutting rate tends to increase as the wood's density and moisture content decreases but appears to be independent of the grain direction. One problem that is often encountered is charring of the cut edge during cutting. However, the extent of charring is usually very small, of the order of microns in depth, and reduces with increasing cutting speed.

Paper and Cardboard Laser cutting of paper and cardboard has an advantage over conventional cutting, since it does not result in broken fibers which may remain attached to the paper and interfere with printing.

Example 9.2 A 1.0 kW CO_2 laser is used to cut a 4.5 mm thick titanium alloy plate using an inert gas assist. The resulting cutting speed and kerf width are 3.0 m/min and 400 μm, respectively.

(a) Calculate the surface temperature at a point 4.0 mm behind the laser beam and 3.5 mm to one side, if the heat transfer efficiency is 90%, and the ambient temperature is 25 °C.
(b) Estimate the size of the HAZ for this process if a cold-rolled alloy recrystallizes at a temperature equal to half the absolute melting temperature of pure titanium.
(c) Obtain the surface temperature for the point in (a) if oxygen assist was used, and 50% of the exothermic reaction contributed to the cutting process.

Solution:
Again, we apply the discussion in Chapter 4 to this problem. The material properties (approximated to be the same as for pure titanium) are obtained from Appendices 4.D and 4.E as

Average density, $\rho = 4507\ \text{kg/m}^3 = 4.507 \times 10^{-6}\ \text{kg/mm}^3$.
Average specific heat, $c_p = 99.3\ \text{J/kg K}$.
Thermal conductivity, $k = 11.4\ \text{W/m K} = 0.0114\ \text{W/mm K}$.
Melting temperature, $T_\text{m} = 1668\ °\text{C} = 1941\ \text{K}$.

Other data:

Atomic weight of titanium is 47.9 g/mol.
Cutting speed is 3.0 m/min = 50 mm/s.

The amount of material, M_t, removed per unit time is

$$M_\text{t} = \rho \times V = 4.507 \times 10^{-6} \times 50 \times 4.5 \times 0.4 = 405.6 \times 10^{-6}\ \text{kg/s} = 0.4056\ \text{g/s}$$

Now the laser power, q_1, available at the workpiece is

$$q_1 = 1000 \times 0.9 = 900 \text{ W}$$

When oxygen assist is used, additional heat available from the exothermic reaction (Eq. (9.15)) is

$$q_2 = 0.5 \times 912 \text{ kJ/mol} = 0.5 \times \frac{912 \times 10^3}{47.9} \times 0.405 \text{ J/s} = 3855.5 \text{ W}$$

Thus, the total power available for cutting is

$$q = q_1 + q_2 = 900 + 3855.5 = 4755.5 \text{ W}$$

Since in cutting, the heat source goes all the way through the thickness, it is considered to be a line source, and thus the heat flow is two-dimensional. The distance, r, of the point of interest from the heat source is thus given by:

$$r = \sqrt{\xi^2 + y^2} = \sqrt{(-4)^2 + 3.5^2} = 5.315 \text{ mm}$$

The thermal diffusivity, κ, of the material is

$$\kappa = \frac{k}{\rho c_\text{p}} = \frac{0.0114}{4.507 \times 10^{-6} \times 99.3} = 25.47 \text{ mm}^2/\text{s}$$

(a) For the temperature at the point (−4, 3.5), we use Eq. (4.22) for a thin plate:

$$T - T_0 = \frac{q}{2\pi kh} \exp\left(-\frac{u_x \xi}{2\kappa}\right) \times K_0\left(\frac{u_x r}{2\kappa}\right)$$

$$\Rightarrow T - 25 = \frac{900}{2\pi \times 0.0114 \times 4.5} \exp\left(-\frac{50 \times (-4)}{2 \times 25.47}\right) \times K_0\left(\frac{50 \times 5.315}{2 \times 25.47}\right)$$

$$\exp\left(-\frac{u_x \xi}{2\kappa}\right) = \exp\left(-\frac{50 \times (-4)}{2 \times 25.47}\right) = 50.71$$

And from Appendix 4.G,

$$K_0\left(\frac{u_x r}{2\kappa}\right) = K_0\left(\frac{50 \times 5.315}{2 \times 25.47}\right) = K_0(5.217) = 0.002902$$

Thus,

$$T = 436\,^\circ\text{C}.$$

(b) For the HAZ size, we use Eq. (4.24) for a thin plate, and since the recrystallization temperature is half the absolute melting temperature, the peak temperature to be used in the equation is 697.5 °C:

$$\frac{1}{T_\text{p} - T_0} = \frac{\sqrt{2\pi e}\rho c_\text{p} h u_x Y}{q} + \frac{1}{T_\text{m} - T_0}$$

$$\Rightarrow \frac{1}{697.5 - 25} = \frac{\sqrt{2\pi \times 2.71828} \times 4.507 \times 10^{-6} \times 99.3 \times 4.5 \times 50 \times Y}{900} + \frac{1}{1668 - 25}$$

And the HAZ size is

$$Y = 1.9 \text{ mm}.$$

(c) The heat input is now 4755.5 kW. Thus,

$$T - 25 = \frac{4755.5}{2\pi \times 0.0114 \times 4.5} \exp\left(-\frac{50 \times (-4)}{2 \times 25.47}\right) \times K_0\left(\frac{50 \times 5.315}{2 \times 25.47}\right)$$

or

$$T = 2196\,^{\circ}\text{C}.$$

This is higher than the melting point of titanium. That means the kerf width will be greater than 7 mm.

9.1.7 Advantages and Disadvantages of Laser Cutting

9.1.7.1 Advantages The general advantages of laser cutting in comparison with conventional cutting processes such as plasma arc cutting, electrical discharge machining (EDM), oxy-acetylene flame cutting, and mechanical cutting are the following:

1. It results in a narrow kerf width, thereby reducing waste. The reduced kerf width also makes the production of arbitrary contours more feasible.
2. Relatively high cutting speeds.
3. Small HAZ due to the relatively small total heat input. Thus, there is very little damage to the base material, making it suitable for heat-sensitive and burnable materials. There is also very little residual stress and distortion.
4. It is a noncontact process, and thus, there is no tool wear and no mechanical forces that could damage delicate workpieces. The lack of mechanical forces also implies less complex fixtures to hold the workpiece in place.
5. Good for both very soft (highly deformable) materials such as paper and very hard (difficult to cut) materials such as diamond.
6. High degree of flexibility (which may facilitate the cutting of complex geometries) and low level of noise.
7. It results in cut edges which are square, and not rounded, as occurs in many other thermal cutting methods.

9.1.7.2 Disadvantages The principal disadvantages of laser cutting are the following:

1. Highly reflective and conductive materials such as gold and silver are difficult to cut using lasers. Methods for improving the absorptivity of metals in general for the laser beam are discussed in Section 11.1.2.3, while those for facilitating cutting in highly reflective materials are discussed in Section 8.1.4.
2. The melting and rapid quenching associated with the process result in a hard edge of the cut piece for hardenable materials.
3. Laser cutting has traditionally been limited to cutting thin materials, that is, less than a few millimeters thick. Furthermore, it is generally limited to cutting through the materials. Thus, blind slots, pockets, or holes are difficult to cut accurately using a laser.

4. The initial capital cost of a laser cutting system is relatively high, about two orders of magnitude higher than that of an oxy-fuel system.
5. Processing of certain materials (such as polymers) may result in the production of dangerous exhaust fumes.

9.1.8 Specific Comparison with Conventional Processes

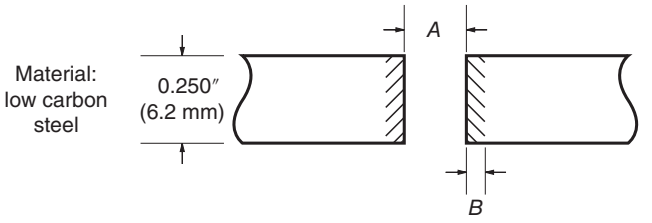
In Sections 9.1.8.1–9.1.8.4, laser cutting is compared with specific alternate processes. Figure 9.20 starts with a numerical comparison of the laser, oxy-acetylene, and plasma arc cutting processes. This is followed in Figure 9.21 with a graphical comparison of EDM, laser cutting, and plasma arc cutting with respect to kerf size, accuracy, maximum cutting speed, and HAZ size. Additional comparison is then made with abrasive waterjet machining and punching/nibbling.

9.1.8.1 Laser, Plasma Arc, and Oxy-Acetylene (Oxy-Fuel) Cutting Each of these three processes uses heat to melt and remove material. Two attributes that are compared in Figure 9.21 for these processes are the kerf width and HAZ size. For both attributes, laser cutting is most advantageous, offering the least kerf width and HAZ size.

9.1.8.2 Laser Cutting and Electrical Discharge Machining (EDM) Unlike the previous case, each of these two processes has certain advantages over the other (Figure 9.21). EDM results in better accuracy and smaller kerf width and HAZ size than laser cutting. Furthermore, EDM is able to cut through relatively thick material. On the other hand, laser cutting is performed at much higher speeds. In addition, the accuracy, kerf width, and HAZ size advantages of EDM are only slight.

9.1.8.3 Laser Cutting and Abrasive Waterjet Machining Abrasive waterjet machining involves cutting through a material using high-pressure water jet. Thus, it does not produce a HAZ. It is also able to cut much thicker materials than does laser cutting. However, the kerf width produced in waterjet machining is less accurate than that associated with laser cutting. Furthermore, waterjet machining is a contact process, and as such, may cause workpiece deflection for weaker materials.

Waterjet machining is often recommended as the process of choice in cutting composites, mainly due to the thermal effect of laser cutting.



	Kerf (*A*)	HAZ (*B*)
Laser	0.012″ (0.3 mm)	0.002″ (0.05 mm)
Oxy-Fuel	0.035″ (0.9 mm)	0.025″ (0.6 mm)
Plasma Arc	0.125″ (3.2 mm)	0.016″ (0.4 mm)

Figure 9.20 Comparison of thermal cutting processes.

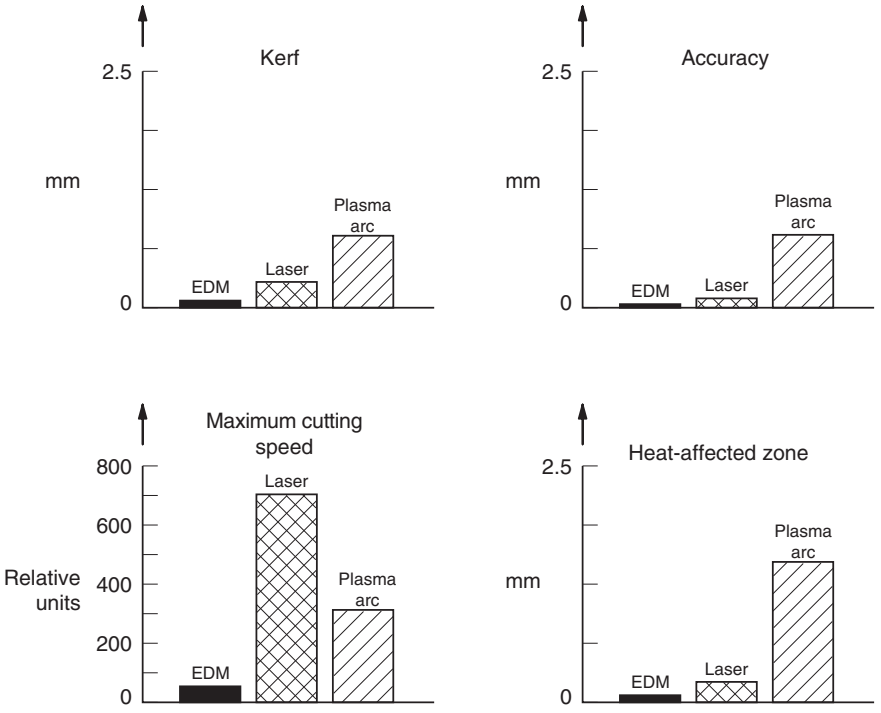


Figure 9.21 Comparison of electrical discharge machining (EDM), laser cutting, and plasma arc cutting processes.

9.1.8.4 Laser Cutting and Punching/Nibbling The primary advantage of laser cutting over punching or nibbling is that it is more flexible, since it does not require tooling. Laser cutting has a definite advantage for small production runs, while at higher production volumes, it loses its advantage, since traditional punching operations are then much faster.

9.1.9 Special Techniques

One special technique that has been developed to enhance the capability of laser cutting is the laser-assisted oxygen cutting process (Lasox). Conventional laser cutting is limited to thicknesses of up to 25 mm, which can be achieved with a laser power of about 3 kW in steels. As discussed in Section 9.1.5.1, the roughness of the cut surface increases as the workpiece thickness increases, as a result of striations which are formed in the thicker material.

The Lasox process is designed to significantly increase the thicknesses that can be cut with a laser of given power without deterioration in quality. The principle is similar to traditional laser cutting and is illustrated schematically in Figure 9.22.

In this process, the laser beam preheats the surface of the workpiece to a temperature close to its melting temperature (>1000 °C for steels), and this facilitates melting of the workpiece by an oxygen gas jet which flows coaxially with the beam. Once the reaction is initiated on the surface, it continues through the workpiece thickness. The cutting action is driven primarily by the gas jet. The footprint of the laser beam is made to cover an area which is just slightly larger than the footprint of the oxygen jet. This is achieved by using a short focal length lens to maximize beam divergence on leaving the nozzle, and yet maintain reasonable stand-off distances. As an illustration, 20–50 mm

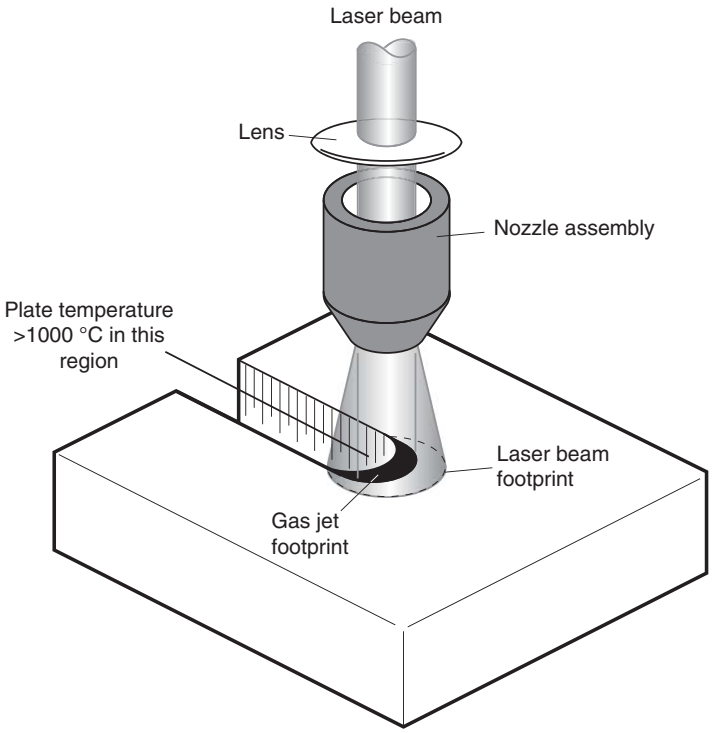


Figure 9.22 Schematic of the laser-assisted oxygen cutting process. Source: Adapted from O'Neill and Gabzdyl (2000).

steel plates can be cut with laser power levels of 700–1100 W at speeds of 0.15–0.5 m/min if the beam diameter on the workpiece surface is 4 mm, gas jet diameter is 3 mm, nozzle exit diameter is 2.5 mm, gas pressure is about 8 bar, and the beam absorption coefficient is 30%.

The surface quality obtained under these conditions is much better than that obtained by either oxy-acetylene cutting alone or by traditional laser cutting, which would require much higher power. The taper that results is also less than 1°, and the top edge is square.

9.2 LASER DRILLING

Laser drilling is used in applications such as jet engine turbine airfoils, injector nozzles, and watches. The process involves applying a laser beam to heat up the material to its melting point or vaporization temperature. When vaporization occurs, it generates a keyhole that results in increased absorptivity, further increasing the hole depth. The molten material or vapor formed is blown away using an assist gas. When the vapor or molten material is ejected out into the surrounding atmosphere, some of it may condense on the workpiece surface as spatter. Very small diameter holes (of the order of microns) can be drilled using a laser, resulting in high aspect ratios.

Our discussion on laser drilling starts with an outline of the different ways in which a laser may be used to drill a hole, followed by a discussion of the basic process parameters. An analysis of the process is then presented, first mentioning the governing equations that will need to be used for a rigorous analysis, followed by an approximate analysis that enables the drilling velocity to be estimated. The advantages and disadvantages of laser drilling are then presented, and finally, potential applications.

9.2.1 Forms of Laser Drilling

There are three ways in which a hole can be produced by laser drilling (Figure 9.23):

1. Single-pulse (or on center) drilling.
2. Multi-pulse percussion drilling.
3. Trepanning.

9.2.1.1 Single-Pulse Drilling This form of drilling involves using a single pulse of laser beam, with a pulse duration of about 1 ms, and energy of several Joules, resulting in a peak power of

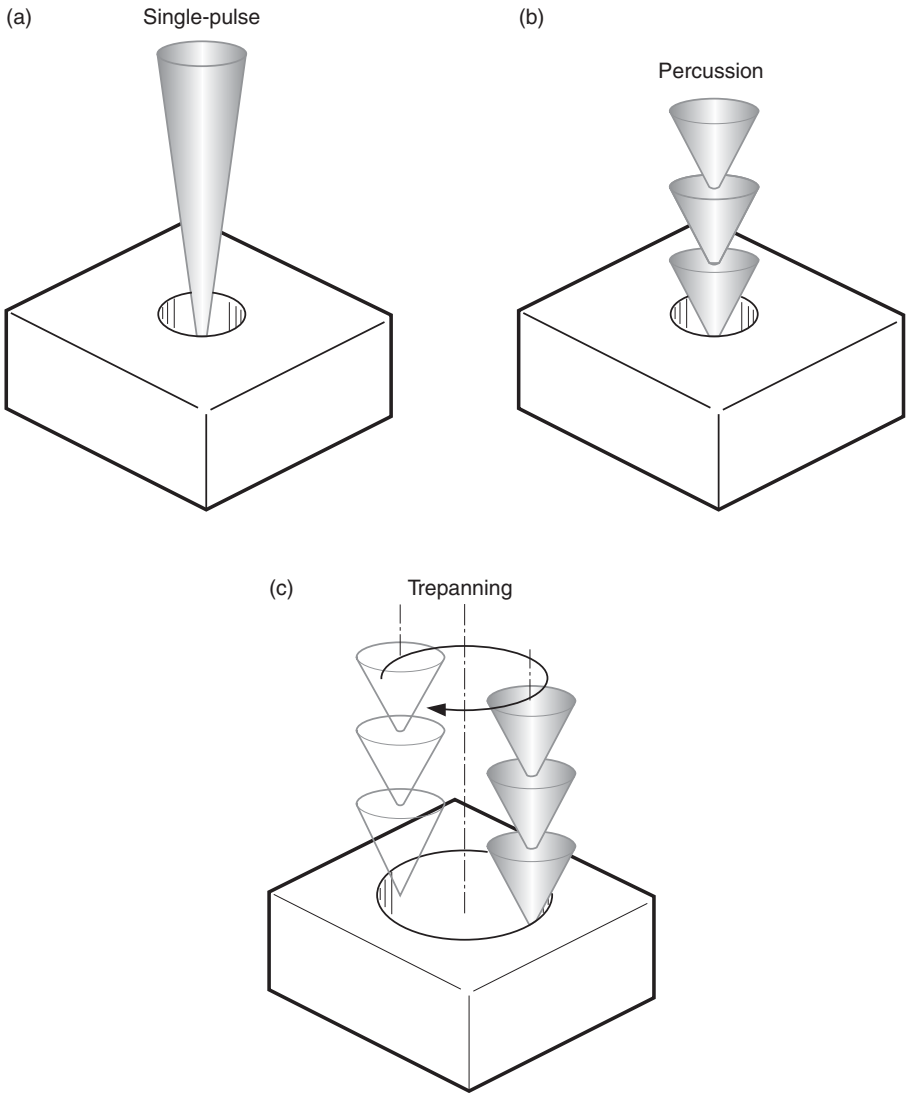


Figure 9.23 Different ways in which a hole may be produced by laser drilling. (a) Single-pulse drilling. (b) Multi-pulse percussion drilling. (c) Trepanning.

the order of 10–100 kW. The material in contact with the laser beam is partly melted and partly evaporated and blown out backward using a gas jet to form the hole. The exceedingly high energy required for drilling very deep holes may result in poor tolerances, reduce the pulse frequency obtainable with the laser, thereby reducing productivity, and increase plasma formation. Plasma formation may affect the hole quality as the laser beam is scattered and/or absorbed and re-radiated to the walls.

9.2.1.2 Multi-Pulse Percussion Drilling Multi-pulse percussion drilling involves drilling a hole using a series of pulses. The hole quality is improved (for example, reduced tapering along the hole length) since the peak power required is reduced, and with it, the amount of molten material and plasma produced during any single-pulse period. However, the series of pulses used reduces the production rate. Typical pulse duration ranges from about 100 fs to 2 ms.

In both percussion modes, the molten and/or evaporated material is ejected out of the hole by the excess pressure of the gas or plasma in the hole over ambient, and this increases with the beam intensity. With increasing hole depth, the pressure required to blow out the molten material increases accordingly. One disadvantage of percussion drilling is that the material that is removed is blown backward toward the lens.

9.2.1.3 Trepanning Trepanning involves cutting out the hole by performing a relative motion (circular or noncircular) between the laser beam and the workpiece. The process is essentially one of laser cutting. Thus, large diameter holes can be produced, with improved hole quality and repeatability. Another advantage of the process is that the material that is removed is blown away from the lens. However, the process is time-consuming, typically requiring about 1 second for producing a hole. Holes drilled by this method are often limited to depths less than 12 mm.

9.2.2 Process Parameters

The type of laser selected for a given application is determined largely by the drilling technique used, the material, and geometry of the hole to be drilled. Nd:YAG lasers are more extensively used in drilling metals than CO_2 lasers when thicknesses exceed 0.5 mm, since absorption of the shorter wavelength 1.06 μm energy by the plasma formed is relatively small, compared to absorption at the 10.6 μm wavelength of the CO_2 laser. Furthermore, higher peak pulse powers are attainable with the Nd:YAG laser. However, both the capital and running costs of CO_2 lasers are lower than those of the Nd:YAG lasers of equivalent power.

The principal parameters associated with percussion laser drilling may be broadly categorized as the beam characteristics (inputs), process characteristics (outputs), and process defects. These are further discussed in the following paragraphs.

9.2.2.1 Beam Characteristics The principal beam characteristics that influence the drilling process include pulse energy, pulse duration, number of pulses, and beam quality. A typical value of the power density used would be 5×10^7 W/cm^2 at a peak power of 8 kW, using a lens of focal length 100 mm.

Pulse Energy: The hole depth generally increases with increasing pulse energy. At much higher energy levels, however, there tends to be greater deformation at the top surface.

Pulse Duration: Laser drilling pulses typically fall in the range of 100 fs to 2 ms. Higher quality holes (for example, reduced taper and distortion) are obtained with shorter beam pulses. However, the number of pulses required to drill a given hole then increases.

Number of Pulses: Generally, hole taper can be reduced in holes with a high depth-to-diameter ratio by using a relatively large number of pulses, while for low depth-to-diameter holes, fewer pulses result in a better quality hole.

Beam Quality: Beams of lower M^2 values or higher quality can be focused to a smaller diameter and also have a longer depth of focus. This increases penetration depth for drilling and cutting, while improving hole quality (high depth-to-diameter ratio with reduced taper), especially for small diameter holes.

9.2.2.2 Drilling Characteristics The principal characteristics of the laser drilling process include the hole diameter, depth, and drilling angle.

Hole Diameter: Typical diameters of laser-drilled holes range in size from 1 μm to 1.5 mm. Hole sizes smaller than 1 μm have been produced using femtosecond lasers. Maintaining a reasonable depth of focus for holes smaller than 12.7 μm is difficult with traditional lasers. For holes of diameter greater than 1.5 mm, it might be more appropriate to use trepanning, which is a relatively slower process.

Hole Depth: In using Nd:YAG lasers, depth-to-diameter ratios as high as 30 : 1 can be achieved.

Drilling Angle: This is the angle that the hole or beam axis makes with the workpiece surface. The power density incident on the workpiece surface is maximum when the workpiece is normal to the beam axis. This configuration also permits the widest range of focal lengths and focusing angles. The drawback, however, is that it results in any reflected beam being redirected back into the optical system. Furthermore, material expelled during the process may be deposited on the lens. This problem may be mitigated by the use of a coaxial gas flow or protective transparent cover plate for the lens.

Increasing the angle between the hole axis and the normal to the workpiece surface reduces the amount of reflected beam and debris that is directed at the lens. However, short focal length lenses are then more difficult to use, and so are gas nozzles. Furthermore, since the beam is then spread over a wider area, the power density is reduced.

9.2.2.3 Process Defects The product quality is affected by defects such as taper of the hole walls, recast, and microcracking.

Taper: Erosion of molten and vaporized material from the hole, and the conical shape of the focused beam, together result in tapering of the drilled hole. The extent of taper is influenced by the pulse energy, number of pulses, and optical system design. It can be minimized by using long focal length lenses, which tend to have longer depth of focus. Tapering is usually not a problem for very small hole depths, say, less than 0.25 mm.

Recast: Recast is the result of resolidified excess molten or vaporized material that is not completely removed. It essentially remains on the newly created surface.

Microcracking: Microcracking normally arises when drilling brittle or high hardenability materials and often results from high cooling rates or temperature gradients.

9.2.3 Analysis of Material Removal During Drilling

As indicated in Section 9.1, when a pulsed beam of adequate intensity is incident on a substrate, the absorbed energy may melt and/or vaporize the substrate material. These are the mechanisms by which drilling takes place. Depending on the energy source and the substrate material, this may occur

by photochemical ablation or photothermal ablation. Photochemical ablation is briefly discussed in Section 9.1.1.3. In this section, we focus on photothermal ablation, which may occur by normal vaporization, normal boiling, or phase explosion.

Normal Vaporization: it refers to the transformation from the condensed phase (solid or liquid) to the vapor phase as atoms or molecules are emitted from the extreme outer surface. It can occur at any fluence and pulse duration, and there is no temperature threshold. The process does not involve nucleation, and the surface temperature is not constant since the vapor pressure is not zero. The contribution of normal vaporization to ablation is insignificant at timescales shorter than 1 ns and also for very low temperatures. The rate at which the drilled surface recedes during normal vaporization, that is, the drilling velocity, u_d, can be obtained from the Hertz–Knudsen equation as:

$$u_d = \left.\frac{\partial z}{\partial t}\right|_{z=0} = \frac{k_c P_s {\lambda_a}^3}{\sqrt{(2\pi m_m k_B T)}} \approx \frac{k_c P_a {\lambda_a}^3}{\sqrt{(2\pi m_m k_B T)}} \exp\left[\frac{L_v m_m}{k_B}\left(\frac{1}{T_v} - \frac{1}{T}\right)\right] \tag{9.17}$$

where k_c is the condensation (or vaporization) coefficient, L_v is the latent heat of vaporization (J/kg), m_m is the mass of the atom or molecule (particle)(kg), P_a is the ambient gas pressure (boiling pressure) (Pa), P_s is the saturated (equilibrium) vapor pressure (Pa), T_v is the vaporization (boiling) temperature corresponding to P_a, with the assumption that there is no vapor present in the ambient and no recondensation (K), and λ_a is the mean atomic spacing of the target (μm).

Normal Boiling: it requires a relatively long pulse period and involves the nucleation of heterogeneous vapor bubbles which may form at the outer surface of the liquid, in the bulk of the liquid, or at the interface between the liquid and surrounding solid. The normal boiling process occurs within the absorption depth ($1/\alpha$, where α is the absorption coefficient). The surface temperature is constant and the same as the vaporization temperature corresponding to the pressure at the surface. The temperature gradient at the surface and also directly beneath it is zero (that is, $\frac{\partial T}{\partial z} = 0$).

Phase Explosion: (also known as explosive boiling). For this to occur, the laser fluence has to be sufficiently high and the pulse duration sufficiently short such that the temperature of the surface and the region immediately beneath it reaches about 90% of the thermodynamic critical temperature (that is, $0.90T_{ct}$). That results in homogeneous bubble nucleation, and the material undergoes a rapid transition from a superheated liquid to a vapor/liquid droplet mixture. Homogeneous nucleation is feasible, since the rate at which it occurs increases significantly near the critical temperature. Here too, the temperature gradient at the surface and also directly beneath it is zero.

Now we focus again on normal vaporization where the substrate material is first heated through the melting point to the vaporization temperature. In this case, part of the beam energy is used to melt the substrate, and the molten material (liquid layer) formed is in direct contact with the substrate. Part of this molten material is further vaporized by the beam. The vapor produced creates a vapor or recoil pressure which pushes the vapor away from the target and also exerts a force on the molten material and expels it sideways.

Material removal may thus occur in both liquid and vapor forms. In the liquid form, the thickness of the liquid layer is estimated to be of the order of 10 μm and decreases with increasing power.

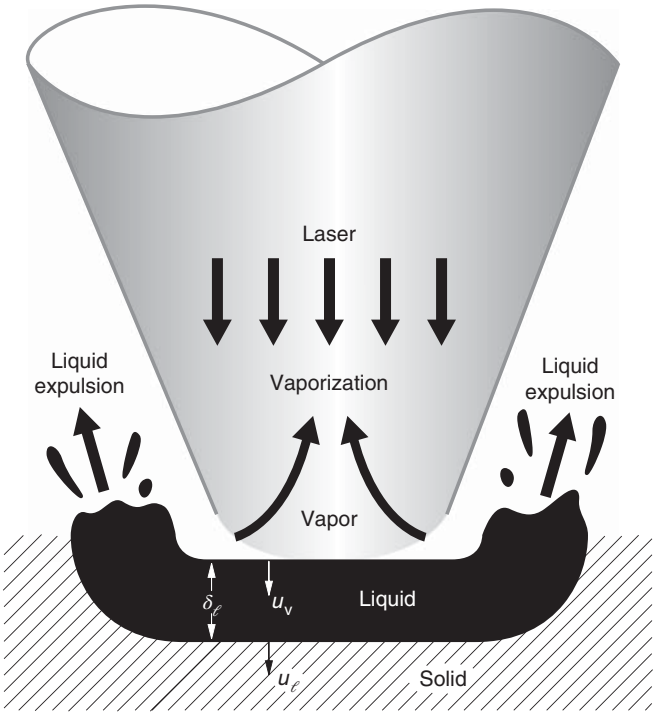


Figure 9.24 Schematic of material removal by a pulsed beam. Source: Adapted from Chan and Mazumder (1987).

In the vapor form, the vaporization rate increases with an increase in power. However, the liquid expulsion rate increases to a maximum and then decreases. At lower power, the vapor temperature is low. Consequently, the expulsion rate is low because of low recoil pressure. At high power, the vapor temperature is high. On the other hand, the thickness of the liquid layer is then so small that the liquid expulsion rate is also reduced. Generally, at low beam power, liquid expulsion is the predominant form of material removal, while vaporization is the predominant form at high power. The process of liquid expulsion is illustrated in Figure 9.24. For beams of much greater intensity and shorter pulse duration, the resulting recoil pressure may be so high as to induce shock waves which are transmitted through the material. Shock waves may also be generated as a result of the interaction of the laser with the plasma.

In the next section, a fundamental approach for analyzing the laser drilling process is outlined. Since the governing equations on which this approach is based are nonlinear, they can only be fully solved using numerical methods. Therefore, in the subsequent section, we present an approximate method for obtaining a quick estimate of the rate at which laser drilling takes place.

9.2.3.1 Basic Analysis The governing equations necessary for analyzing the laser drilling process are the basic continuity, momentum, and energy equations, with the appropriate boundary conditions. There are two interfaces associated with the material removal process. One is the solid–liquid interface which can be treated as a conventional Stefan problem, Section 6.1.1. The other is the liquid–vapor interface, adjacent to which may exist a Knudsen layer with a thickness of a few molecular mean free paths. Across this layer, continuum theories are no longer valid since discontinuities

exist in temperature, density, and pressure. The discontinuity across the Knudsen layer is the basis for the shock wave.

9.2.3.2 Approximate Analysis In this section, we present a simple extension of the Hertz–Knudsen equation for estimating the rate (velocity) of the drilling process, which takes into account the contribution of the ejected molten material. The following assumptions are made:

1. Heat conduction and vapor absorption losses are negligible.
2. The thermal properties of the material are constant, and thermal expansion is negligible.
3. The skin depth ($\approx 10^{-5}$ mm) (Section 8.1.6) is negligible.

If we consider the energy flux carried away with the expulsed material to be equal to the irradiance or power density (power per unit area) absorbed by the material, I_a, then we can write

$$I_a = \varphi_v \Delta H_v + \varphi_l \Delta H_l \tag{9.18}$$

where ΔH is the specific energy absorbed by the expulsed material (J/kg), ϕ is the expulsion rate (kg/m^2 s), and the subscripts v and l refer to the vapor and liquid, respectively.

The drilling velocity, u_d, can be expressed in terms of ϕ_v and ϕ_l as:

$$u_d = \frac{1}{\rho}(\varphi_v + \varphi_l) \tag{9.19}$$

In developing an expression for ϕ_l, one may consider a simple one-dimensional conduction heat flow for a system with a constant moving heat source. The quasi-static temperature distribution, T, in the medium for this case, neglecting any phase changes and lateral heat conduction losses, may be approximated as:

$$T = \frac{I'_a}{\rho c_p u_d} \exp\left[-\frac{\xi u_d}{\kappa}\right] \tag{9.20}$$

where $I'_a = I_a - \varphi_v L_v$ is the difference between the absorbed power density and that carried away from the absorbing layer by the evaporated material (W/m^2), L_v is the latent heat of vaporization (J/kg), u_d is the drilling velocity (m/s), κ is the thermal diffusivity of the material (m^2/s), and ξ is the moving coordinate pointing in the direction of u_d.

The rate of evaporation, ϕ_v, of a hot surface at a temperature T_h is obtained from the Hertz–Knudsen equation as:

$$\varphi_v = (1 - R) P_s \left(\frac{m_m}{2\pi k_B T_h}\right)^{1/2} \tag{9.21}$$

where k_B is Boltzmann's constant (J/K), m_m is the mass of the atom or molecule (particle) (kg), and $R \approx 0.2$ is the mean particle reflection coefficient of the metal surface.

The saturated vapor pressure, P_s, may be approximated by:

$$P_s = P_a \exp\left[\frac{L_p}{k_B T_v}\left(1 - \frac{T_v}{T_h}\right)\right] \tag{9.22}$$

where L_p is the heat of evaporation per particle (J/particle), P_a is the ambient pressure (Pa), T_h is the hot surface temperature (K), and T_v is the vaporization temperature (K).

The hot surface temperature, T_h, may be estimated as the temperature at $\xi = 0$ in Eq. (9.23). Then again considering heat balance, but only for the molten material, we have

$$\int_{T_0}^{T_h} c_p dT + L_m = \frac{I'_a}{\rho u_d} \tag{9.23}$$

where L_m is the latent heat of melting (J/kg) and $T_0 = T_a$ is the ambient temperature (K).

The thickness, δ_l, of the liquid layer formed by heat conduction may be estimated using Eq. (9.23) as:

$$\delta_l = \left(\frac{\kappa}{u_d}\right) \ln\left[\frac{T_h}{T_m}\right] \tag{9.24}$$

where T_m is the melting temperature.

The pressure, P, of the evaporating surface acting on the liquid layer pushes the liquid radially outward with a velocity, u_l, determined using Bernoulli's Eq. (9.1a) as:

$$u_l = \left(\frac{2P}{\rho}\right)^{1/2} \tag{9.25}$$

The expelled liquid escapes through an opening (Figure 9.24) that is determined by the beam radius, w, of an area A_e given by:

$$A_e = 2\pi w \delta_l \tag{9.26}$$

Since the actual pressure P is about half the saturation pressure P_s (Eq. (9.22)), the liquid expulsion rate, ϕ_l, with respect to the irradiated surface is obtained from Eqs. (9.24)–(9.26) as:

$$\varphi_l = \left[\left(\frac{2\kappa}{w}\right) \ln\left(\frac{T_h}{T_m}\right)\right]^{1/2} {P_s}^{1/4} \rho^{3/4} \tag{9.27}$$

Equations (9.21) and (9.27) give the expulsion rates of the vapor and liquid in terms of the surface temperature, T_h. However, T_h, in turn, depends on ϕ_l and ϕ_v, (Eq. (9.23)). Thus, actual calculation of the expulsion rates will need to be done iteratively. However, for a first approximation, the surface temperature, T_h, may be estimated from Eq. (9.20) and assumed constant.

It must be noted that the drilling speed calculated by this approach refers to the average rate at which material is removed during the laser pulse. It does not give the overall production rate.

Example 9.3 An Nd:YAG laser generates pulses of intensity 60 MW/cm^2 for drilling 200 μm diameter holes in a tungsten plate. Estimate the drilling speed that can be achieved for this operation. Assume ambient temperature and pressure of 25 °C and 0.1 MPa, respectively, and that the diameter of the hole is the same as the beam diameter.

Solution:
Some of the material properties are obtained from Appendices 4.D and 4.E as

Average density, $\rho = 19254\ \text{kg/m}^3 = 19.254 \times 10^{-6}\ \text{kg/mm}^3$
Average specific heat, $c_p = 134\ \text{J/kg K}$
Thermal conductivity, $k = 151.2\ \text{W/m K} = 0.1512\ \text{W/mm K}$
Latent heat of fusion, $L_m = 220\ \text{kJ/kg} = 220{,}000\ \text{J/kg}$
Latent heat of vaporization, $L_v = 4815\ \text{kJ/kg} = 4815 \times 10^3\ \text{J/kg}$

Melting temperature, $T_m = 3410\,^\circ\text{C} = 3683\,\text{K}$
Evaporation temperature, $T_v = 5700\,^\circ\text{C} = 5973\,\text{K}$.

Other needed properties are as follows:

Boltzmann's constant, $k_B = 1.38 \times 10^{-23}$ J/K
Avogadro's number, $N_0 = 6.023 \times 10^{23}$ molecules
Atomic weight, $W_a = 183.8$ g/mol $= 0.1838$ kg/mol.

Calculated properties:

Particle mass, $m_p = 0.1838\,\text{kg}/6.023 \times 10^{23} = 3.05 \times 10^{-25}$ kg/particle
Heat of vaporization per particle, $L_p = 4815 \times 10^3\,\text{J/kg} \times 3.05 \times 10^{-25} = 14.69 \times 10^{-19}$ J/particle
Incident power density is given by $I_a = 60\,\text{MW/cm}^2 = 60 \times 10^4\,\text{W/mm}^2$
Thermal diffusivity, $\kappa = \frac{k}{\rho c_p} = \frac{0.1512}{19.254 \times 10^{-6} \times 134} = 58.6\,\text{mm}^2/\text{s} = 58.6 \times 10^{-6}\,\text{m}^2/\text{s}$.

We assume an initial value for the drilling speed of 2.5×10^4 mm/s $= 25$ m/s, which is used to estimate the hot surface temperature using Eq. (9.20):

$$T - T_0 = \frac{I_a'}{\rho c_p u_d} \exp\left[-\frac{\xi u_d}{\kappa}\right]$$

but

$$I_a' = I_a - \varphi_v L_v$$

Since ϕ_v is not known, we initially use $I_a' = I_a$ to estimate the hot surface temperature.

From Eq. (9.23), assuming a constant specific heat, the hot surface temperature can be expressed as:

$$T_h = T_0 + \frac{I_a}{\rho c_p u_d} - \frac{L_m}{c_p}$$

or

$$T_h = 25 + \frac{60 \times 10^4}{19.254 \times 10^{-6} \times 134 \times 2.5 \times 10^4} - \frac{22 \times 10^4}{134}$$
$$\Rightarrow T_h = 7686\,^\circ\text{C} = 7959\,\text{K}$$

Now from Eq. (9.22), the saturation pressure is given by:

$$P_s = P_a \exp\left[\frac{L_p}{k_B T_v}\left(1 - \frac{T_v}{T_h}\right)\right]$$

Now $P_a = 10^5\,\text{N/m}^2 = 0.1\,\text{N/mm}^2$. Thus,

$$P_s = 0.1 \times \exp\left[\frac{14.69 \times 10^{-19}}{1.38 \times 10^{-23} \times 5973}\left(1 - \frac{5973}{7959}\right)\right]$$
$$= 8.537\,\text{MPa}\,(\text{N/mm}^2) = 8.537 \times 10^6\,\text{Pa}\,(\text{N/m}^2)$$

Therefore, from Eq. (9.21),

$$\begin{aligned}\varphi_v &= (1-R)P_s\left(\frac{m_m}{2\pi k_B T_h}\right)^{1/2}\\ &= (1-0.2)\times 8.537\times 10^6\left(\frac{3.05\times 10^{-25}}{2\pi\times 1.38\times 10^{-23}\times 7959}\right)^{1/2}\\ &= 4540\,\text{kg/m}^2\text{s}\end{aligned}$$

And from Eq. (9.27)

$$\varphi_l = \left[\left(\frac{2\kappa}{w}\right)\ln\left(\frac{T_h}{T_m}\right)\right]^{1/2} P_s{}^{1/4}\rho^{3/4}$$

$$\begin{aligned}\varphi_l &= \left[\left(\frac{2\times 58.6\times 10^{-6}}{0.100\times 10^{-3}}\right)\ln\left(\frac{7959}{3683}\right)\right]^{1/2}[8.537\times 10^6]^{1/4}\times[19{,}254]^{3/4}\\ &= 8.4\times 10^4\,\text{kg/m}^2\text{s}\end{aligned}$$

This indicates that under these conditions, melting is the predominant mode of material removal. Now from Eq. (9.19), the drilling velocity is given by:

$$\begin{aligned}u_d &= \frac{1}{\rho}(\varphi_v+\varphi_l)\\ &= \frac{1}{19{,}254}(84{,}000+4540)\\ &= 4.6\,\text{m/s} = 4.6\times 10^3\,\text{mm/s}\end{aligned}$$

This is about an order of magnitude lower than the original estimate. Thus, an iteration is necessary. The calculated value for ϕ_v can be used to update the value of I_a' during the iteration process.

9.2.4 Advantages and Disadvantages of Laser Drilling

9.2.4.1 Advantages The advantages that laser drilling has over conventional drilling are the following:

1. It is a noncontact process. There are no machining forces, and thus no deflection of a tool and/or workpiece that would result in dimensional errors in the part produced, or wear of the tool.
2. Accurate location of holes. In conventional drilling, the drill often deflects on coming in contact with the workpiece, resulting in drill "wandering." This affects accuracy of hole location and orientation.
3. There is no chip formation, and thus no need for elaborate chip disposal systems. However, an appropriate exhaust system is often necessary for the vaporized material.
4. Laser drilling results in high depth-to-diameter aspect ratios, but primarily for holes of relatively small diameter, and depths of a few millimeters.
5. Ability to drill difficult-to-machine materials such as diamond, ceramics, and highly refractory metals, on which conventional drilling performs poorly.

6. Ability to drill holes with difficult entrance angles. Drilling of holes with low entrance angles is quite difficult with conventional drills. However, with laser drilling, entrance angles as low as 10° are possible.
7. Ability to drill very small holes. Small diameter conventional drills are liable to break easily. On the other hand, the lower end of hole sizes that can be drilled using a laser is limited by the beam size at the focal point, with hole sizes of the order of microns being achievable.
8. High degree of flexibility, enabling a number of operations to be combined in a single setup.
9. High duty cycles and high production rates. Since lasers can be used continuously without interruption, high duty cycles are achieved, resulting in relatively high production rates.

9.2.4.2 Disadvantages The primary limitations of the laser drilling process include the following:

1. Inability to produce precision blind holes. Lasers are more suited to producing through holes.
2. Inability to drill deep holes. The hole depths that are achieved by laser drilling are relatively small, being typically a few millimeters deep. As the hole depth increases, the diameter begins to enlarge due to beam expansion, resulting in a tapered hole.
3. Holes that are close to an edge, and those that intersect other holes are often not effectively produced by laser drilling.

9.2.5 Applications

Typical applications for holes drilled using Nd:YAG lasers include holes in jet engine turbine airfoils and combustors (hole sizes about 180 μm–1 mm), injector nozzles, and oil passages in automotive engine or gear components. Since CO_2 lasers are more effectively used on nonmetals, typical drilling applications involving CO_2 lasers include sound suppression holes in aircraft engine liners made of polymers or polymeric composites, drilling holes in ceramic substrates in the microelectronic industry, and cigarette paper perforating.

9.3 NEW DEVELOPMENTS

9.3.1 Micromachining

Micromachining involves machining applications that result in feature sizes on the order of tens of micrometers or less. As indicated in the section on laser drilling, hole diameters of the order of a few microns can be produced using traditional lasers, with pulse durations ranging from CW to hundreds of picoseconds. Common lasers used for this purpose include the CO_2, Nd:YAG, fiber, copper vapor, and excimer lasers. However, the mechanism of material removal with such long-pulse lasers involves both melting and evaporation (Figure 9.24). This results in significant dross formation (Section 9.1.5.2), which affects hole quality. Excimer lasers, being the exception, result in direct photochemical ablation (Section 9.1.1.3) due to the ultraviolet wavelength of the beam, and that improves the hole quality.

When a laser beam is incident on a material, the photon energy is first absorbed by electrons which then get heated to high temperatures rapidly. The electron energy gets transferred to the lattice through electron–phonon collisions (Section 4.2.5), resulting in heating of the lattice. The extent of lattice heating depends on the pulse duration and the electron–phonon coupling. For long pulse durations, the electrons and lattice can reach thermal equilibrium, and a relatively large portion of the lattice gets heated by diffusion, with a resulting peak temperature that is relatively low. The material is then first heated through the melting temperature to the vaporization temperature (if the power

is high enough). A relatively large volume around the laser focus melts, but only a relatively small layer of material reaches the vaporization temperature.

As the pulse duration decreases, the laser field intensity increases, and the interaction time gets shorter. Thus, for much shorter pulse durations, the intensity is much higher, resulting in relatively high electron temperatures (of the order of tens of electron volts) during the pulse. Most of the electron–phonon coupling takes place after the pulse and often results in lattice temperatures that are higher than are achieved with long pulses for the same total energy. Most of the heating is confined to the interaction zone and passes rapidly through the melting temperature to the vaporization temperature. Thus, a large fraction of the material in the interaction volume is vaporized, going through a melt phase very quickly. With ultrashort pulses, when the laser fluence just exceeds the threshold value, the region that is vaporized can be smaller than the focal area (Figure 9.25). The reproducible fabrication resolution is generally limited to about 100 nm as a result of pulse to pulse energy fluctuations and beam output variations.

Ultrashort-pulse lasers ($\leq$1 ps) with wavelengths in the visible and near-infrared range therefore result in high precision micromachining, with clean and uniform holes or cuts. The breakdown threshold of such lasers can be determined with high accuracy since it is deterministic. Since the beam intensity profile can be controlled such that only a specified portion of the focused beam spot is above the threshold (Figure 9.25), a region of diameter smaller (1 μm or less) than the focused beam size (5 μm or less) can be ablated, with barely any heat being conducted out of the processing zone. The mechanics of laser–material interaction that results in ablation under such conditions are discussed in the following paragraphs.

Absorption of a laser beam that is incident on a material occurs by linear or nonlinear processes. The absorption results in heating of material in the focal region first to the melting temperature, and then to the vaporization temperature, depending on the beam intensity and pulse duration. For radiation outside the ultraviolet range, the absorption mechanisms are different for materials such as metals and semiconductors which are absorbing, and glasses and plastics which are transparent dielectric materials. Absorption also depends on laser intensity. For opaque materials, absorption is linear when the beam pulse duration is long and the intensity low, while nonlinear absorption becomes dominant when the pulse duration is ultrashort and the intensity high.

9.3.1.1 Transparent Dielectric Materials Absorption for transparent materials results mainly from nonlinear processes. Avalanche ionization and multiphoton ionization are the nonlinear mechanisms that cause breakdown.

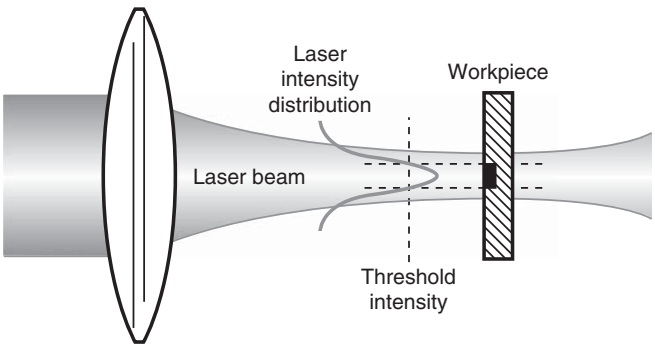


Figure 9.25 Setting the breakdown threshold to a specific level of beam mode. Source: Adapted from Liu et al. (1997).

Avalanche Ionization The bound valence electrons of transparent dielectric materials have an ionization potential or bandgap which is greater than the laser photon energy (Section 2.3.1). Therefore at low beam intensities (typical of long-pulse lasers), the laser energy is not absorbed by the bound electrons. However, a few free electrons (seed electrons) that exist in the material as a result of impurities do absorb laser energy when they collide with bound electrons and lattice (inverse bremsstrahlung process). Acceleration of the seed electrons provide them with kinetic energy that is in excess of the ionization potential of the bound electron. Thus as these seed electrons collide with bound electrons, ionization occurs, resulting in additional free electrons, in much the same way that a photon interacting with an excited atom results in another photon being generated. Repetition of this process leads to an avalanche that increases the free-electron density exponentially. At a critical density (10^{18}/cm^3 for pulse durations greater than a nanosecond), the transparent material breaks down and becomes absorbing. The absorption at the critical density is significant enough that irreversible damage takes place. At the critical density, the plasma oscillation frequency equals the laser frequency, and the transparent material becomes totally opaque. For ultrashort pulses, the critical density, ρ_e, is given by:

$$\rho_e = \frac{m_e \omega^2}{4\pi e^2} \tag{9.28}$$

where m_e is electron mass and ω is the laser frequency.

The critical electron density is then about 10^{21}/cm^3 for lasers in the visible and near-infrared wavelength range. The breakdown process typically results in the generation of acoustic waves and radiation of optical plasma or spark at the focus.

Multiphoton Ionization At very high beam intensities, bound electrons of transparent materials are directly ionized (that is, lifted from the valence band to the conduction band or free energy level) by multiphoton absorption. This is normally the case with ultrashort-pulse beams and involves simultaneous absorption of m photons during the pulse such that

$$mh_p v \geq E_g \tag{9.29}$$

where E_g is the bandgap energy or ionization potential.

Generally for the critical density to be achieved that results in irreversible breakdown and therefore ablation, the laser fluence must exceed a threshold value, which depends on the pulse duration. For pulse durations longer than a few tens of picoseconds, the threshold fluence is proportional to the square root of the pulse duration. Furthermore, there is significant scatter associated with the threshold fluence, making it more stochastic in nature. Avalanche ionization is the principal mechanism for breakdown in this pulse duration range. For ultrashort pulses, the seed electrons are created by multiphoton ionization, but even though the actual breakdown process is primarily accomplished through avalanche ionization, the threshold is more deterministic or precise.

9.3.1.2 Metals and Semiconductors Materials which are absorbing (such as metals and semiconductors) generally have a significant number of free and valence electrons whose ionization potential (bandgap) is smaller than the laser photon energy. Linear absorption is therefore the predominant mode when the laser pulse duration is long, resulting in Joule heating. Here too breakdown, which results in melting and vaporization, is accompanied by the generation of acoustic waves and radiation of optical plasma or spark at the focus.

There are two characteristic lengths associated with absorption of the laser beam. One is the skin depth, δ', (Section 8.1.6) which is the distance within which the photon energy is absorbed and may be expressed as:

$$\delta' = \frac{1}{\beta} \tag{9.30}$$

where β is the decay constant.

The other is the heat diffusion length, l_h, which is the distance over which the thermal energy is conducted during the pulse and can be expressed as:

$$l_h = \sqrt{\kappa \tau_p} \tag{9.31}$$

where τ_p is the pulse duration.

For long pulses, the fluence threshold varies as the square root of the pulse duration since $l_h > \delta'$, and therefore the volume of material heated by the beam pulse depends on the heat diffusion length. There is then significant scatter associated with the threshold. Since the volume of material that is heated is relatively large, significant amount of melting occurs, and during ablation, material is expelled in both liquid and vapor form by the recoil (vapor) pressure associated with the ablation process (Section 9.2.3). The unexpelled molten material resolidifies around the ablated region, resulting in dross formation (Figure 9.26a). The thickness of melt layer that results for long-pulse lasers can be as much as 10 times that of the evaporation layer. That means the feature sizes that can be obtained are relatively large.

When the pulse duration becomes short enough that $l_h < \delta'$, then the threshold becomes independent of the pulse duration and is therefore more deterministic. Furthermore, since the energy absorbed is highly localized, the material in this region is quickly heated through the melting temperature to the vaporization temperature. Most of the ablation process then occurs by direct vaporization, resulting in more precise and cleaner cuts (Figure 9.26b). A significant portion of the material is removed after the laser pulse has ended. Even though heat diffusion results in heating of the

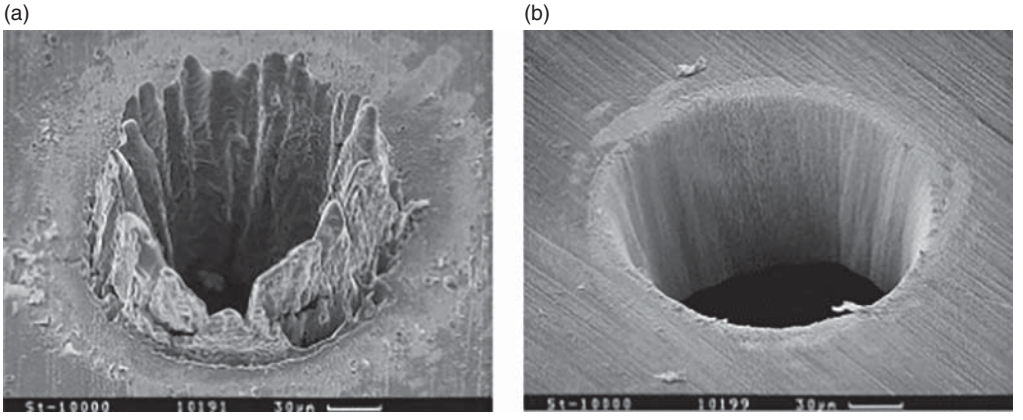


Figure 9.26 SEM photographs of holes drilled through a 100 μm steel foil using Ti:sapphire laser pulses at 780 nm. (a) Pulse duration = 3.3 ns and fluence = 4.2 J/cm^2. (b) Pulse duration = 200 fs and fluence = 0.5 J/cm^2. Source: Momma et al. (1997), Elsevier.

surrounding material, the amount of material that melts as a result is relatively small. This is due in part to the rapid cooling that follows the heating process, and also to the significant amount of energy that is lost by vaporization. With ultrashort-pulse lasers, the melt thickness may only be about twice that of the evaporation layer. Thus, submicron feature sizes are achievable.

The dependence of the fluence threshold on the square root of the pulse duration when $l_h > \delta'$ also applies to transparent materials.

Advantages of ultrashort-pulse lasers in relation to material ablation can therefore be summarized as the following:

1. The precision of the ablation (breakdown) threshold (which depends on the pulse duration, being higher for the longer pulse duration beam).
2. Reduction in the laser fluence needed to initiate ablation.
3. Significant reduction in HAZ size.

In recent years, ultrashort-pulse lasers have found increasing use in machining microchannels, such as those used in microfluidics and photovoltaic cells. They have also been used to drill holes of diameter 35 μm in 1 mm thick stainless steel sheet, resulting in good quality holes with straight sides. Hole sizes as small as 0.3 μm have been produced in silver films using 200 fs pulses produced by a Ti:sapphire laser which was focused to a spot size of 3 μm. As with traditional lasers, femtosecond lasers can be used in machining materials ranging from diamond to biodegradable polymers.

9.3.1.3 Micro-Explosions At very high intensities, material removal can also occur by the process of micro-explosion, which triggers shock and rarefaction waves, thereby creating voxels that result in voids within the material. This provides a means for scribing or creating internal geometrical structures in solid transparent dielectrics such as glasses, polymers, and crystals. The voxel diameters are as small as 0.2–0.25 μm. The micro-explosion may result from multiphoton absorption, which is initiated in a highly localized focal volume of a tightly focused beam where the intensity is high enough (in excess of 10^{14} W/cm^2). Thus, the scribing can occur within the interior of the transparent material where the beam is focused, at depths of about 50–200 μm. For this to be effective, a high numerical aperture (about 1.35) lens is necessary. It may also occur as a result of individual femtosecond pulses that create optical breakdown, resulting in extreme temperature and pressure conditions.

The observed voxel size is often smaller than what would be expected from a region with intensity level above the necessary threshold intensity (Figure 9.27). The smaller size is due to self-focusing as a result of intensity dependence of the refractive index:

$$n(r) = n_0 + n_2 I(r) \tag{9.32}$$

where $I(r)$ is the intensity at a radial distance r, $n(r)$ is the refractive index at r, n_0 is the nominal refractive index, n_2 is proportional to the third-order susceptibility, and thus requires very high intensities.

Self-focusing can be avoided by ensuring that the laser power used is below the threshold power level (q_{th}) for self-focusing, which is given by:

$$q_{th} = \frac{\lambda^2}{2\pi n_0 n_2} \tag{9.33}$$

By eliminating self-focusing, the beam energy can be delivered to a precise location.

Femtosecond lasers (typically 100 fs) are most suited for this process due to the ultrashort duration and high intensity of the pulses. In the focal region, the material is first transformed into solid density plasma (electron density of about $10^{23}/cm^3$) with a skin depth of about 60 nm. The plasma absorbs about 60% of the incident energy, resulting in a pressure of the order of 10 TPa. A shock wave is thus initiated at the boundary of the region where the energy is deposited, with a corresponding rarefaction wave propagating inward to create the void. As the shock wave propagates in the cold material, it loses energy. It becomes converted into a sound wave when the energy inside the shock-affected volume becomes equal to the absorbed energy. That defines the boundary of the shock-affected volume. At that point, the pressure behind the shock wave then equals the internal pressure of the cold material. The sound wave propagates into the rest of the material without causing any permanent structural changes. Material in the unfocused region is unaffected by the beam. Example conditions for silica would be a 30–100 nJ pulse energy at 0.8 μm wavelength, and 180 fs duration. With a single laser pulse in a sapphire crystal, the void formed is surrounded by an amorphous sapphire shell of diameter approximately twice that of the void. With longer pulse lasers (200 ps or 10 ns), voxels produced tend to have irregular shapes and are larger in size. They also usually have cracks that extend radially.

Formation of a void in a confined region is only possible if that is accompanied by compaction of the surrounding material, since the material removed to create the void has no room to escape. The intense pressure of the explosion then creates the denser surrounding material (shock-affected region) which is more reactive chemically and thus etches differently from the surrounding material. The explosion pressure is the result of a temperature rise at constant volume. For crystalline sapphire and high purity silica, the radius, r_s, of the shock-affected region may be approximated by:

$$r_s = \sqrt[3]{Q_{abs}/\frac{4}{3}\pi E_m} \tag{9.34}$$

where Q_{abs} is the energy absorbed and E_m is Young's modulus.

The equation is based on the micro-explosion process only and does not account for self-focusing or void refill by back-relaxation of the shock-displaced material. Thus, the final voxel size may deviate from what is predicted by the equation.

One potential application of a process involving micro-explosion is three-dimensional data storage. The technique is also applicable to semiconductors provided the frequency of the laser is below the bandgap.

9.3.1.4 Micromachining Applications Micromachining has application in several industries including the automotive (diesel fuel injectors), biomedical (stent (Section 14.1.1)), aerospace (aero engines), and electronic (circuit board hole drilling) industries, as well as surface texturing (Section 11.6) for many other industries.

9.3.2 Laser-Assisted Machining

Laser-assisted machining is a means for increasing productivity in metal removal for traditionally difficult-to-machine materials such as titanium alloys, nickel-based superalloys, hardened steels, ceramics, and metal matrix composites. The metal removal rate and productivity achieved when machining such materials by conventional methods are limited by the significant reduction in tool life at high speeds.

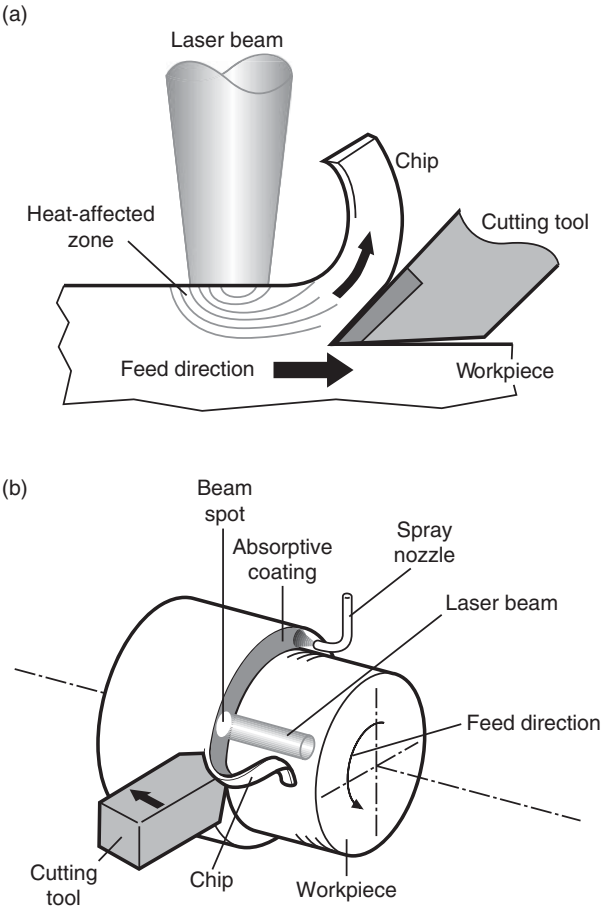


Figure 9.27 Schematic of the laser-assisted machining process. (a) Two-dimensional view. (b) Three-dimensional view. Source: From Weck and Kasperowski (1997)/MIT Publishing Ltd.

The process involves using a laser beam to heat up a portion of a material directly in front of a cutting tool to a temperature lower than its melting point, but high enough to soften the material ahead of the cutting tool (Figure 9.27). Increasing the temperature of a material normally decreases its flow stress and strain hardening rate. Ordinarily, it should be possible to reduce the flow stress simply by machining at high speeds. This is feasible with materials such as aluminum alloys whose flow stress decrease rapidly with increasing temperature. However, many of the difficult-to-machine materials tend to retain their strength at elevated temperatures. Thus, high-speed machining of such materials increases tool failure rate. In conventional machining, the chip is hottest at the underside which is in contact with the cutting tool, while the opposite is true of laser-assisted machining, even though in the latter case, the temperature of the underside of the chip may be the same as that obtained in conventional machining.

Laser-assisted machining is essentially an improvement of the process of using a gas torch or induction heating to facilitate the machining process. Gas torch and induction heating are not commercially acceptable because of the resulting large HAZ and distortion of the machine and

workpiece. The high heat input also affects the cutting tool. Some of the benefits of laser-assisted machining include

1. Reduction in cutting forces.
2. Improvement in tool life.
3. Improvement in surface finish.

Reduction in cutting forces increases the accuracy of cutting by decreasing the amount of workpiece and machine distortion. It also minimizes the tendency for tool fracture in interrupted cuts, especially when ceramic cutting tools are used. Looking at it from another viewpoint, by keeping the cutting forces constant, higher feed rates can be used, enabling higher material removal rates to be achieved. Improvement in surface finish of the workpiece improves its fatigue life as well as wear and corrosion resistance.

To minimize the effect of the applied heat on the cutting tool, the depth of heat penetration into the shear zone has to be limited to a region above the cutting tool. By so doing, the workpiece is softened without overheating the cutting tool.

In laser-assisted turning of 316 and 420 stainless steels and the nickel superalloy, udimet 700, using a pulsed laser beam of power 460 W, with a maximum average power density of 2.3 MW/cm^2, it has been found that the use of laser assist produces a smoother workpiece surface. Even though the 420 stainless steel does not show a reduction in cutting forces, the improvement in surface finish obtained is significant. For the 316 stainless steel and 1018, 1040 plain carbon steels, force reductions in the range of 25–50% can be expected. For 1090 steel, much higher forces (about 2–3 times) are obtained with the laser on than when it is off. This is most likely due to hardening of the workpiece by the laser beam through phase transformation. An improvement in surface quality, though, is obtained. It has been found that laser-assisted machining in general is not effective at very high machining speeds, since the interaction time at an irradiated spot is too short for the metal to be heated at the high speeds. At very low speeds, on the other hand, the material cools between the time that it is irradiated and the time it arrives at the tool. Hardenable materials thus experience higher forces at low speeds. For a previously hardened material, the cutting forces are always reduced. For either type of material, the forces are reduced at intermediate speeds that result in the maximum amount of tempering of the material arriving at the tool.

Conventional machining of titanium alloys is normally done at relatively low speeds and exclusively with carbide cutting tools, due to the relative weakness of ceramics and the high reactivity of titanium. Also, the low thermal conductivity of titanium alloys tends to concentrate any heat generated in the immediate vicinity of the tool, accelerating tool failure. In laser-assisted machining of titanium alloys, it has been found that effective results are obtained with the beam spot oriented such that its minor axis is along the direction of rotation (Figure 9.27b). The amount of time that any point on the workpiece is impinged by the beam is then minimal, thereby reducing heat penetration below the shear plane and into the cutting tool. This is especially true when carbide tools are used, since they are less resistant to high temperatures than ceramic tools. Due to the low speeds, absorptive coatings may not be necessary. At a laser power level of 12 kW, cutting speed of 0.7 m/s, feed rate 0.38–2.00 mm/rev, and depth of cut 2.5 mm, productivity is doubled by increasing the feed rate if the cutting forces and tool wear are kept constant. The low conductivity of titanium alloys results in a reduced heat penetration, thereby reducing the impact of the laser beam on tool wear. Since titanium alloys are normally machined at low speeds, much greater improvements can be expected with the application of laser-assisted machining.

In laser-assisted machining of inconel 718 using ceramic tools with a laser power of 12 kW, cutting speed of 2.0 m/s, feed rate 0.15 mm/rev, and depth of cut 1.3 mm, it has been found that impinging the beam at the top edge of the shoulder substantially reduces notch wear at the depth

of cut line but has little influence on nose wear. Nose wear could be reduced by focusing the beam near the root of the shoulder, but then that results in no improvement in the notch wear. In either case, however, flank wear is reduced. The forces have been found to decrease almost linearly with increasing laser power, with the rate being a little over 1% per kilowatt when no absorptive coatings are used on the workpiece surface. With absorptive coatings, the reductions average nearly 4% per kilowatt at power levels up to 4 kW. However, at power levels greater than 4 kW, the force curves level out, indicating a progressive decrease in absorptivity. At a speed of 0.5 m/s and power 14 kW, a 33% reduction in force can be obtained without a coating, but the forces may be reduced by an additional 33% when a potassium silicate coating is used. Overall, it can be deduced that metal removal rates could be increased by one third in laser-assisted machining, while at the same time reducing both tool wear and cutting forces.

Coating techniques are discussed in greater detail in Section 11.1.2.3. For laser-assisted machining, the necessary criteria to be satisfied by successful coatings include

1. Good dispersion with fine particle size to avoid intermittent disruption of spray.
2. Localized application.
3. Rapid bonding to workpiece surface.
4. Good thermal stability under the beam, with minimal volatilization or steaming.
5. High absorptivity.
6. Lack of abrasiveness.
7. Inertness to workpiece and ease of removal.

Process Parameters:
The principal parameters that affect the process quality are the following:

1. Beam power and size (power density).
2. Distance between the beam and cutting tool. If the beam impinges on the workpiece very far from the tool, the thermal front can penetrate through to the tool and aggravate, rather than alleviate tool wear.
3. Cutting speed.
4. Depth of cut.

In turning operations, the beam may be positioned as shown in Figure 9.27a. To minimize the penetration of heat below the shear plane, the distance between the beam and cutting tool has to be as small as possible. However, the actual distance depends on processing parameters such as the beam intensity, cutting speed, and depth of cut. The technique has also been applied to other machining operations such as milling, drilling, and grinding.

9.4 SUMMARY

Laser cutting and drilling both involve using a focused laser beam to rapidly heat up a material to the melting and/or vaporization temperature, and blowing away the molten material/vapor using an assist gas. The main difference is that in laser cutting, the ejected material is blown away from the laser optics, whereas in laser drilling, it is partly blown toward the optics.

The three principal forms of laser cutting are fusion cutting, sublimation cutting, and photochemical ablation. A laser cutting system in its simplest form consists of the laser generator and beam delivery system, a nozzle assembly for the assist gas, a motion unit, and an exhaust for the waste

material. The key parameters that influence the process are the beam power and characteristics, traverse speed, assist gas, and relative location of the focal point. As an example, a 3 mm thick stainless steel plate can be cut with a CO_2 laser using an output power of 1.8 kW at a speed of 2 m/min and an 800 kPa nitrogen assist gas pressure. As the material thickness increases, the cutting speed for a given power decreases. Common assist gases used are oxygen, air, inert gases, and nitrogen. Oxygen and air result in an exothermic reaction that may increase the cutting speed, depending on the material.

A Gaussian beam is generally preferred for cutting, since it results in the highest power density at the focal point. Even though beams which are plane-polarized in a plane parallel to the cutting direction result in higher cutting speeds than beams which are plane-polarized normal to the cutting direction, circularly polarized beams are generally used since they result in better cut quality when the cutting direction changes. It is also generally preferred to have the focal point either on the workpiece surface or just below it.

An analysis of the process shows that as the cutting velocity increases, the thickness of molten material at the cutting front increases, but on the other hand, increasing the gas jet velocity decreases the melt thickness. The quality of cut that is obtained is often determined by striations and dross that form on the cut surface. Thicker workpieces result in more irregular striations, culminating in a rougher cut surface. Dross formation is enhanced by higher surface tension or viscosity of the molten material. Thus, using oxygen assist tends to reduce dross formation, as opposed to when inert gas assist is used.

Lasers can be used to produce holes in parts in three ways: percussion drilling, which involves using a single laser pulse; multi-pulse percussion drilling, where a series of pulses are used; and trepanning, which actually makes a circular cut to create the hole. The drilling process is influenced primarily by the pulse energy, pulse duration, number of pulses, and beam quality. Holes as small as 1 μm can be made, and depth-to-diameter ratios as high as 30 : 1 can be achieved. Common defects that affect product quality are tapering of the hole walls, recast, and microcracking. Typical applications of laser drilling include holes in jet engine turbine airfoils and injector nozzles.

As the pulse duration of the laser beam decreases, so does the extent to which heat is diffused into the surrounding material, for the same input energy. Furthermore, the material is heated more rapidly through the melting range. Thus, cleaner cuts are obtained with ultrashort-pulse lasers, compared to traditional long-pulse lasers. The ablation threshold of ultrashort-pulse lasers is also deterministic, enabling better control of the process.

Material removal can also occur by micro-explosion, a process that results from multiphoton absorption when the beam intensity is very high as a result of tight focusing. This triggers shock and rarefaction waves, creating voids within the material. It therefore provides a means for scribing or creating internal geometrical structures in solid transparent dielectrics such as glasses, polymers, and crystals.

Combining a laser with traditional machining processes can reduce cutting forces, while also improving surface finish and tool wear.

APPENDIX 9.A

List of symbols used in the chapter

Symbol	Parameter	Units
Subscripts g, l, v	Refer to gas, liquid, and vapor, respectively	—
A_e	Area of opening through which expelled liquid escapes	mm^2

Symbol	Parameter	Units
h_x	Melt thickness	μm
H_l	Specific energy absorbed by expelled liquid	kJ/kg
H_v	Specific energy absorbed by expelled vapor	kJ/kg
I_a	Irradiance absorbed by material	kW/m^2
k_c	Constant	—
l_h	Heat diffusion length	m
L_p	Heat of evaporation per particle	J/particle
m_m	Mass of atom or molecule (particle)	g
P_s	Saturated vapor pressure	Pa
r_c	Local radius of curvature	mm
R	≈0.2 = mean particle reflection coefficient	—
T_{ct}	Thermodynamic critical temperature	K
T_h	Temperature of a hot surface	K
T_l	Liquid temperature	K
T_m	Melting temperature	K
T_v	Evaporation temperature	K
u	Velocity	m/s
u_c	Cutting velocity	m/s
u_d	Drilling velocity	m/s
u_{g0}	Gas velocity at nozzle (reference)	m/s
u_i	Velocity components (u_x, u_y, u_z)	m/s
z_i	Height of location	mm
β	Decay constant	μm^{-1}
δ'	Skin depth	μm
δ_l	Thickness of liquid layer formed by heat conduction	μm
ΔH	Specific energy absorbed by the expulsed material	J/kg
ΔH_{vm}	Molar enthalpy of vaporization	J/mol
γ_l	Surface tension of the molten material	N/m
λ_a	Mean atomic spacing of target	μm
φ	Expulsion rate	kg/m^2 s
τ	Shear stress distribution at the cutting front	Pa
τ_{gl}	Shear stress at interface between gas and molten metal	Pa
τ_p	Pulse duration	s
θ_i	Inclination of the cut surface with the vertical	°

PROBLEMS

9.1. Given the quasi-static temperature distribution, T, as:

$$T = \frac{I'_a}{\rho c_p u_d} \exp\left[-\frac{\xi u_d}{\kappa}\right]$$

Show that the thickness, δ_l, of the liquid layer formed by heat conduction may be expressed as:

$$\delta_l = \left(\frac{\kappa}{u_d}\right) \ln\left[\frac{T_h}{T_m}\right]$$

9.2. Starting with Bernoulli's equation, show that the liquid layer is pushed radially outward with a velocity, u_1, given by:

$$u_1 = \left(\frac{2P}{\rho}\right)^{1/2}$$

9.3. A 5 mm thick plain carbon steel plate is cut using a 2.5 kW laser at a speed of 4.5 m/min.
(a) Calculate the surface temperature at a point 4 mm behind the laser beam and 1.5 mm to one side, if the heat transfer efficiency is 80%, and the ambient temperature is 30 °C.
(b) Estimate the size of the HAZ for this process.

9.4. A 15 × 10 mm rectangular piece is to be cut out of a polymer sheet of thickness 8 mm, using a laser. The output laser power available at the workpiece is 1.0 kW, and the part is to be cut at a speed of 2.0 m/min, resulting in a kerf width of 0.3 mm. If the reflectivity, R, of the polymer sheet is 25%, and another 4.0% of the beam is lost and not used for cutting, determine the specific energy (energy per unit volume) of the material.

9.5. (a) Which form of laser cutting involves melting of the base material which is then ejected using assist gas?
(b) Which form of laser cutting uses the least energy?
(c) How does orientation of the plane of polarization of the laser beam relative to the cutting direction affect the quality of the cut surface?
(d) A plot of cutting speed versus workpiece thickness usually has two limiting curves. What happens when one selects speeds which are below the lower limiting curve?

9.6. A 5 mm thick nickel plate is cut using a 4 kW CO_2 laser at a speed of 1.5 m/min. If the resulting kerf width is 0.45 mm, the ambient temperature is 25 °C, and 15% of the incident laser beam is absorbed by the plate, determine
(a) The amount of energy that is supplied by the laser while cutting a circular blank of diameter 0.5 m.
(b) The amount of material (in kg) that is removed or lost as a result of the process.
(c) The energy required to heat the material that is removed, to the melting point.
(d) The energy required to melt the material after it is heated to the melting point.
(e) The energy that is lost by conduction, convection, and radiation.

9.7. Estimate the drilling speed that can be expected while drilling 350 μm diameter holes in a low carbon steel plate using a 1.5 kW CO_2 laser. Assume ambient temperature and pressure of 25 °C and 0.1 MPa, respectively.

9.8. If the kerf width that results when a steel plate of thickness 2 mm is cut with a 2 kW CO_2 laser is 0.25 mm, determine the cutting speed. Assume the ambient temperature is 25 °C and that 15% of the incident laser beam is absorbed by the plate. You may assume

Average density, $\rho = 7870\ \text{kg/m}^3$

Average specific heat, $c = 452\ \text{J/kg-K}$

Latent heat of fusion, $L_m = 247\ \text{kJ/kg}$

Melting temperature, $T_m = 1538\ °\text{C} = 1811\ \text{K}$

Further assume that all material is removed by melting, and that there is no vaporization.

9.9. (a) Assuming most of the content of a dried human bone is calcium, and further assuming that calcium sublimes (turns from solid into gas directly), determine the drilling speed as a function of temperature of the hot bone surface using the Hertz–Knudsen equation.

Then determine the drilling speed using approximate analysis and discuss the difference between the two methods.

(b) Determine the depth of focus for both Nd:YAG and CO_2 lasers with incident beam radius of 0.2 mm at a focal length of 0.1 m.

(c) Determine the power required for both Nd:YAG and CO_2 under the same conditions and find out which would be preferable in terms of power consumption for the operation.

9.10. In cutting a mild steel plate of thickness 7.5 mm with a 4.5 kW CO_2 laser, the kerf width is determined to be 2.0 mm. If the cutting speed is 12.5 cm/min, determine

(a) The total amount of energy used in cutting out a circular piece of radius 7.5 cm.

(b) The amount of energy used in heating the material to the melting point.

(c) The energy used in actually melting the material once it reaches the melting point.

(d) The percentage of laser power used in melting the material, from room temperature.

(e) The percentage of laser energy that is lost by conduction, convection, and radiation.

You may assume the same properties and ambient conditions as in Problem 9.8.

Further assume that all material is removed by melting, and that there is no vaporization.

BIBLIOGRAPHY

Abeln, T., Radtke, J., and Dausinger, F. (1999). High precision drilling with short-pulsed solid-state lasers. *International Congress on Applications of Lasers & Electro-Optics (*Vol. 1999, No. 2, pp. 195–203). Laser Institute of America.

Atanasov, P.A. and Gendjov, S.I. (1987). Laser cutting of glass tubing – A theoretical model. *Journal of Physics D: Applied Physics* 20: 597–601.

Bass, M., Beck, D., and Copley, S.M. (1978). Laser assisted machining. *SPIE Fourth European Electro-Optics Conference* 164: 233–240.

Batanov, V.A., Bunkin, F.V., Prokhorov, A.M., and Fedorov, V.B. (1973). Evaporation of metallic targets caused by intense optical radiation. *Soviet Physics JETP* 36 (2): 311–322.

Beyer, E. and Petring, D. (1990). State of the art laser cutting with CO_2 lasers. *International Congress on Applications of Lasers & Electro-Optics* 1990 (1): 199–212.

Bronstad, B.M. (1993). Capability testing of laser cutting machines. *Welding in the World* 31: 27–30.

Bunting, K.A. and Cornfield, G. (1975). Towards a general theory of cutting: a relationship between the incident power density and the cut speed. *ASME Journal of Heat Transfer* 97: 116–121.

Canisius, M., Herzog, D., Schmidt-Lehr, M. et al. (2015). Laser cutting of carbon fiber-reinforced plastic with an absorber transparent for visible spectrum. *Journal of Laser Applications* 27: 032003. https://doi.org/10.2351/1.4916532.

Caristan, C. (2004). *Laser Cutting Guide for Manufacturing*. Dearborn, Michigan: Society of Manufacturing Engineers.

Chan, C. and Mazumder, J. (1987). One-dimensional steady-state model for damage by vaporization and liquid expulsion due to laser-material interaction. *Journal of Applied Physics* 62: 4579–4586.

Chen, M.M. (1992). Generics structure of flow and temperature fields in welding and high energy beam processing, *Heat and Mass Transfer in Materials Processing, Hemisphere Publishing Co., New York*, pp. 365–382.

Chen, S.L. and Steen, W.M. (1991). The theoretical investigation of gas assisted laser cutting. *International Congress on Applications of Lasers & Electro-Optics* 1991 (2): 221–230.

Chryssolouris, G. (1991). *Laser Machining, Theory and Practice*. Berlin: Springer-Verlag.

Cronin, M.J. (1992). Metallurgical response of nickel base superalloys to laser machining. *International Congress on Applications of Lasers & Electro-Optics* 1992 (1): 469–479.

Dong, X. and Shin, Y.C. (2016). Multi-scale modeling of alumina ceramics undergoing laser-assisted machining. *ASME Journal of Manufacturing Science and Engineering* 138, 011004-1. https://doi.org/10.1115/1.4029858.

Dunsky, C. and Colville, F. (2008). Solid state laser applications in photovoltaics manufacturing. *Proceedings of SPIE* 6871, 687129-1.

Edler, R. and Berger, P. (1991). New nozzle concept for cutting with high power lasers. *International Congress on Applications of Lasers & Electro-Optics* 1991 (2): 253–262.

Fieret, J., Terry, M.J., and Ward, B.A. (1986). Aerodynamic interactions during laser cutting. *SPIE, Laser Processing: Fundamentals, Applications, and SystemsEngineering* 668: 53–62.

Gabzdyl, J.T. and Morgan, D.A. (1992). Assist gases for laser cutting of steels. *International Congress on Applications of Lasers & Electro-Optics* 1992 (1): 443–448.

Geiger, M., Bergmann, H.W., and Nuss, R. (1988). Laser cutting of sheet steels. *SPIE, Laser Assisted Processing* 1022: 20–33.

Geiger, M. and Gropp, A. (1993). Laser fusion cutting of stainless steel with a pulsed Nd:YAG laser. *Welding in the World* 31: 22–26.

Glezer, E.N. and Mazur, E. (1997). Ultrafast-laser driven micro-explosions in transparent materials. *Applied Physics Letters* 71 (7): 882–884. https://doi.org/10.1063/1.119677.

Hashimoto, T., Juodkazis, S., and Misawa, H. (2007). Void formation in glasses. *New Journal of Physics* 9: 253. https://doi.org/10.1088/1367-2630/9/8/253.

Heglin, L.M. Introduction to laser drilling. In: *The Industrial Laser Annual Handbook*, 116–120. Tulsa: PennWell Books.

Herzog, D., Schmidt-Lehr, M., Canisius, M. et al. (2015). Laser cutting of carbon fiber reinforced plastic using a 30 kW fiber laser. *Journal of Laser Applications* 27: S28001. https://doi.org/10.2351/1.4906304.

Ignatiev, M., Okorokov, L., Smurov, I. et al. (1994). Laser assisted machining: process control based on real-time surface temperature measurements. *Journal de Physique IV* 4: C4-65–C4-68.

Juodkazis, S., Nishimura, K., Tanaka, S. et al. (2006). Laser-induced micro-explosion confined in the bulk of a sapphire crystal: evidence of multi-megabar pressures. *Physical Review Letters, PRL* 96: 166101.

Juodkazis, S., Misawa, H., Hashimoto, T. et al. (2006). Laser induced micro-explosion confined in a bulk of silica: formation of nanovoids. *Applied Physics Letters* 88: 201909. https://doi.org/10.1063/1.2204847.

Juodkazis, S., Misawa, H., Gamaly, E.G. et al. (2009). Is the nano-explosion really microscopic? *Journal of Non-Crystalline Solids* 355: 1160–1162.

Kawasumi, H. (1990). Laser processing in Japan. In: *Industrial Laser Annual Handbook*, 141–143. Tulsa: Penwell Books.

Kelly, R. and Miotello, A. (1996). Comments on explosive mechanisms of laser sputtering. *Applied Surface Science* 96-98: 205–215.

Ki, H. (2001). Modeling and measurement of processes with liquid-vapor interface created by high power density lasers. Ph.D. Thesis. University of Michigan, Ann Arbor.

Kim, C.-J., Kauh, S., Ro, S.T., and Lee, J.S. (1994). Parametric study of the two dimensional keyhole model for high power density welding processes. *Journal of Heat Transfer* 116: 209–214.

Kitagawa, A. and Matsunawa, A. (1990). Three dimensional shaping of ceramics by using co_2 laser and its optimum processing condition. *International Congress on Applications of Lasers & Electro-Optics* 1990 (1): 294–301.

Knight, C.J. (1979). Theoretical modeling of rapid surface vaporization with back pressure. *AIAA Journal* 17: 519–523.

König, W. and Zabokliski, A.K. (1993). Laser assisted hot machining of ceramics and composite materials. *Machining of Advanced Materials*, NIST Special Publication 847: 455–463.

Krokhin, O.N. (1972). Generation of high-temperature vapors and plasmas by laser radiation. In: *Laser Handbook* (ed. F.T. Arecchi and E.O. Schulz-Du Bois), 1371–1407. North-Holland Publishing Co.

Kuger, T., and Naeem, M. (2002). Material processing with super-modulation, *Proceedings of the 21st International Congress on Applications of Lasers and Electro-Optics (ICALEO 2002), Scottsdale, Arizona, USA*.

Landau, L.D. and Lifshitz, E.M. (1969). Phase equilibrium. In: *Statistical Physics*, 261. Oxford: Pergamon Press.

Lei, S., Shin, Y.C., and Incropera, F.C. (2001). Experimental investigation of thermo mechanical characteristics in laser-assisted machining of silicon nitride ceramics. *ASME Journal of Manufacturing Science and Engineering* 123: 639–646.

Lei, S., Zhao, X., Yu, X. et al. (2020). Ultrafast laser applications in manufacturing processes: a state of art review. *ASME Journal of Manufacturing Science and Engineering* 142: 031005. https://doi.org/10.1115/1.4045969.

Li, M., Li, S., Yang, X. et al. (2018). Fiber laser cutting of CFRP laminates with single- and multi-pass strategy: a feasibility study. *Optics and Laser Technology* 107: 443–453.

Liu, X., Du, D., and Mourou, G. (1997). Laser ablation and micromachining with ultrashort laser pulses. *IEEE Journal of Quantum Electronics* 33 (10): 1706–1716.

Marot, G., Fan, L.J., Tarrats, A. et al. (1991). The workpiece material interaction and the laser assisted machining. *Annals of the CIRP* 40: 91–94.

Marot, G., Cohen, P., Salem, W.B. et al. (1992). Reduction in cutting force by laser assisted machining contribution of optimum condition research experiments. *Bulletin du Cercle d'betudes des mbetaux* 16 (4): 1–7.

Masumoto, I., Kutsuna, M., and Ichikawa, K. (1992). Relation between process parameters and cut quality in laser cutting of aluminum alloys. *Transactions of the Japan Welding Society* 23 (2): 7–14.

Miotello, A. and Kelly, R. (1999). Laser-induced phase explosion: new physical problems when a condensed phase approaches the thermodynamic critical temperature. *Applied Physics A - Materials Science & Processing* 69: S67–S73.

Miyamoto, I. and Maruo, H. (1991). The mechanism of laser cutting. *Welding in the World* 29 (9/10): 12–23.

Mizeikis, V., Kimura, S., Surovtsev, N.V. et al. (2009). Formation of amorphous sapphire by a femtosecond laser pulse induced micro-explosion. *Applied Surface Science* 255: 9745–9749.

Molian, P.A. (1993). Dual-beam CO_2 laser cutting of thick metallic materials. *Journal of Materials Science* 20: 1738–1748.

Momma, C., Nolte, S., Chichkov, B. et al. (1997). Precise laser ablation with ultrashort pulses. *Applied Surface Science* 109 (110): 15–19.

Niino, H., Harada, Y., and Fujisaki, A. (2016). Deep-trench structure of carbon-fiber-reinforced thermoplastics (CFRTP) by laser cutting with a single-mode IR fiber laser. *JLMN-Journal of Laser Micro/Nanoengineering* 11 (3): 372–376.

O'Neill, W., Gabzdyl, J.T., and Steen, W.M. (1992). The dynamical behavior of gas jets in laser cutting. *International Congress on Applications of Lasers & Electro-Optics* 1992 (1): 449–458.

O'Neill, W. and Steen, W.M. (1995). A three-dimensional analysis of gas entrainment operating during the laser-cutting process. *Journal of Physics D: Applied Physics* 28: 12–18.

O'Neill, W. and Gabzdyl, J.T. (2000). New developments in laser-assisted oxygen cutting. *Optics and Lasers in Engineering* 34: 355–367.

Olsen, F.O. and Heckermann, S. (1990). High speed laser drilling. *International Congress on Applications of Lasers & Electro-Optics* 1990 (1): 141–150.

Ortiz, A.L. (1990). On-the-fly drilling with a fiber delivered face pumped laser beam. *ICALEO Proceedings* 1990: 151–166.

Ostendorf, A., Kamlage, G., and Chichkov, B.N. (2003). Precise deep drilling of metals by femtosecond laser pulses. *RIKEN Review* (50): 87–89.

Petring, D., Abels, P., and Beyer, E. (1988). Absorption distribution on idealized cutting front geometries and its significance for laser beam cutting. *SPIE, High Power CO_2 Laser Systems and Applications* 1020: 123–131.

Pizzi, P. (1987). Laser metalworking. In: *Applied Laser Tooling* (ed. O.D.D. Soares and M. Perez-Amor), 213–233. Netherlands: Martinus Nijhoff Publishers.

Powell, J., Frass, K., Menzies, I.A., and Fuhr, H. (1988). CO2 laser cutting of non-ferrous metals. *SPIE, High Power CO_2 Laser Systems and Applications* 1020: 156–163.

Powell, J., Jezioro, M., Menzies, I.A., and Scheyvearts, P.F. (1988). CO_2 laser cutting of titanium alloys. *SPIE, Laser Technologies in Industry* 952: 609–617.

Powell, J., Ivarson, A., Kamalu, J. et al. (1992). The role of oxygen purity in laser cutting of mild steel. *ICALEO Proceedings* 1992: 433–442.

Rajagopal, S. (1982). Laser-assisted machining of tough materials. *Laser Focus* 18 (3): 49–54.

Rajagopal, S., Plankenhorn, D.J., and Hill, V.L. (1982). Machining aerospace alloys with the aid of a 15 kw laser. *Journal of Applied Metalworking* 2 (3): 170–184.

Ready, J.F. (1965). Effects due to absorption of laser radiation. *Journal of Applied Physics* 36 (2): 462–468.

Ready, J.F. (ed.) (2001). *LIA Handbook of Laser Materials Processing*. Orlando, FL: Laser Institute of America, Magnolia Publishing, Inc.

Ream, S.L., Bary, P., and Perozek, P.M. (1990). Heat input during high pressure laser cutting of inconel and titanium. *ICALEO Proceedings* 1990: 213–224.

Rizvi, N.H. (2003). Femtosecond laser machining: current status and applications. *RIKEN Review* (50): 107–112.

Schock, W., Giesen, A., Hall, T. et al. (1986). Transverse radio frequency discharge: a promising excitation technique for high power CO_2 lasers. *SPIE, Laser Processing: Fundamentals, Applications, and System Engineering* 668: 246–251.

Schreiner-Mohr, U., Dausinger, F., and Hugel, H. (1991). New aspects of cutting with CO_2 lasers. *ICALEO Proceedings* 1991: 263–272.

Schulz, W., Simon, G., Urbassek, H.M., and Decker, I. (1987). On laser fusion cutting of metals. *Journal of Physics D: Applied Physics* 20: 481–488.

Schulz, W. and Becker, D. (1989). On laser fusion cutting: the self-adjusting cutting kerf width. *SPIE, High Power Lasers and Laser Machining Technologies* 1132: 211–221.

Schulz, W., Becker, D., Franke, J. et al. (1993). Heat conduction losses in laser cutting of metals. *Journal of Physics D: Applied Physics* 26: 1357–1363.

Schuocker, D. (1986). Laser cutting. *SPIE, The Industrial Laser Annual Handbook* 629: 87–107.

Schoucker, D. and Muller, P. (1987). Dynamic effects in laser cutting and formation of periodic striations. *SPIE, High Power Lasers* 801: 258–264.

Schuocker, D. (1988). Heat conduction and mass transfer in laser cutting. *SPIE, Laser Technologies in Industry* 952: 592–599.

Schmidt, A.O. (1949). Hot milling, *Iron Age*, April.

Shin, J.S., Oh, S.Y., Park, H. et al. (2018). Laser cutting of steel plates up to 100 mm in thickness with a 6-kW fiber laser for application to dismantling of nuclear facilities. *Optics and Lasers in Engineering* 100: 98–104.

Shin, Y., Wu, B., Lei, S. et al. (2020). Overview of laser applications in manufacturing and materials processing in recent years. *ASME Journal of Manufacturing Science and Engineering* 142: https://doi.org/10.1115/1.4048397.

Sider, T. and Steffen, J. (1990). YAG-SLAB laser cutting performance. *SPIE, High Powered Solid State Lasers and Applications* 1277: 232–243.

Simon, G. and Gratzke, U. (1989). Theoretical investigations of instabilities in laser gas cutting. *SPIE, High Power Lasers and Laser Machining Technologies* 1132: 204–210.

Takeno, S., Moriyasu, M., and Hiramoto, S. (1992). Laser drilling by high peak pulsed CO_2 laser. *ICALEO Proceedings* 1992: 459–468.

Tiffany, W.B. (1985). Drilling, marking, and application for industrial Nd:YAG lasers. *SPIE, Applications of High Powered Lasers* 527: 28–36.

Tour, S. and Fletcher, L.S. (1949). Hot spot machining, *Iron Age*, July.

Ursu, I., Nistor, L.C., Teodorescu, V.S. et al. (1984). Early oxidation stage of copper during CW CO_2 laser irradiation. *Applied Physics Letters* 44: 188–189.

van Krieken, A.H., Schaarsberg, G.J., and Raterink, H.J. (1988). Laser micro machining of material surfaces. *SPIE, Laser Assisted Processing* 1022: 34–37.

Vicanek, M. and Simon, G. (1987). Momentum and heat transfer of an inert gas jet to the melt in laser cutting. *Journal of Physics D: Applied Physics* 20: 1191–1196.

Vicanek, M., Simon, G., Urbassek, H.M., and Decker, I. (1987). Hydrodynamical instability of melt flow in laser cutting. *Journal of Physics D: Applied Physics* 20: 140–145.

von Allmen, M. (1976). Laser drilling velocity in metals. *Journal of Applied Physics* 47: 5460–5463.

Webber, T. (1991). Laser cutting of galvanized steel: CO_2 vs. CW Nd:YAG. *ICALEO Proceedings* 1991: 273–281.

Weck, M. and Kasperowski, S. (1997). Integration of lasers in machine tools for hot machining. *Production Engineering* IV (1): 35–38.

Yaroslava, G., Yingling, G., and Garrison, B.J. (2003). Photochemical ablation of organic solids. *Nuclear Instruments and Methods in Physics Research B* 202: 188–194.

Yilbas, B.S., Davies, R., and Yilbas, Z. (1992). Study into penetration speed during CO_2 laser cutting of stainless steel. *Optics and Lasers in Engineering* 17: 69–82.

Znotins, T., Poulin, D., and Eisele, P. (1990). Excimer laser: continued growth in materials processing applications. In: *The Industrial Laser Annual Handbook* (ed. D. Belforte and M. Levitt), 102–105. PennWell Publishing Company.

Zweig, A.D. (1991). A thermo-mechanical model for laser ablation. *Journal of Applied Physics* 70: 1684–1691.

10 Laser Welding

Welding is one means for joining parts. Others include mechanical fastening which uses physical devices such as bolts and rivets to hold parts together, and adhesive bonding which uses nonmetallic materials to hold parts together, for example, glueing. Welding, on the other hand, uses heat and/or pressure to produce localized coalescence of material. It may also require filler metal, shielding, or fluxing. Welding processes can be categorized into two broad areas, namely fusion welding processes where a portion of the pieces to be joined actually melts and non-fusion welding processes. Examples of the former include arc welding, resistance welding, gas, laser, and electron beam welding processes, and for the latter, we have brazing, soldering, and solid-state welding processes. Our discussion in this chapter focuses on laser welding.

The most common form of lasers used for welding are CO_2, Nd:YAG, and fiber lasers. However, continuous welds have traditionally been done with CO_2 lasers, since higher-power continuous wave (CW) beams are obtained with CO_2 lasers. Recent developments in high-power CW Nd:YAG and fiber lasers have increased their application in this area.

The discussion in this chapter starts with a general outline of laser welding parameters. This is followed by a discussion on welding efficiency and then the mechanism of the laser welding process. After that, the behavior of different materials, including some common metals and their alloys, ceramic materials, as well as dissimilar materials, when welded, are presented. Common discontinuities encountered during laser welding are then outlined, followed by advantages and disadvantages of the process. Special laser welding techniques, such as multiple-beam welding, arc-augmented laser welding, wobble welding, and remote laser welding (RLW), are then presented. This is followed by a discussion on specific applications such as microwelding and laser-welded tailored blanks. Finally, laser transmission welding and laser brazing are discussed.

10.1 LASER WELDING PARAMETERS

The setup for laser welding is similar to the one used for laser cutting (Figure 9.1). The primary difference is that in laser welding, the exhaust system is positioned above the beam delivery unit rather than the other side of the workpiece. The principal parameters that affect the laser welding process include the following:

1. Beam power and traverse speed.
2. Beam characteristics (mode, stability, polarization, and form – pulsed or CW).
3. Shielding gas.
4. Location of focal point relative to the workpiece surface.
5. Joint configuration.

Principles of Laser Materials Processing: Developments and Applications, Second Edition. Elijah Kannatey-Asibu, Jr.
© 2023 John Wiley & Sons, Inc. Published 2023 by John Wiley & Sons, Inc.
Companion Website: www.wiley.com/go/Kannatey-Asibu/PrinicplesLaserMaterialsProcessing

Table 10.1 Sample conditions for deep penetration laser welding of different materials.

Material	Weld type	Thickness (mm)	Welding speed (m/min)	Laser power (kW)	Laser type
Carbon steel	Butt	12	0.8	9	CO_2
Carbon steel	Butt	6	10	25	CO_2
Stainless steel (304)	Bead on plate	20	0.1	10	Fiber
Stainless steel (304)	Butt	20	1.27	20	CO_2
Aluminum (A5083)	Bead on plate	10	10	10	Fiber
Titanium	Butt	4.5	5.08	16	CO_2

Table 10.2 Typical processing parameters for deep penetration laser welding.

Laser power	1–20 kW
Welding speed	0.5–10 m/min
Shielding gas	He, Ar @ 5–40 L/min
Nozzle diameter	2–4 mm
Nozzle–workpiece distance	10 mm
Depth of focus	Upper 1/3rd in material thickness
Focusing optic	f = 150–1600 mm
Beam mode	TEM_{00}, multimode

As an illustration of conditions that might be used in industry, Table 10.1 gives sample conditions for CO_2 and fiber laser welding of a variety of materials, while Table 10.2 provides additional details for a specific material. We now discuss each of the parameters listed earlier in further detail in the following sections.

10.1.1 Beam Power and Traverse Speed

Figure 10.1 shows the effect of welding speed and laser power on weld penetration. The depth of penetration decreases almost exponentially with increasing traverse speed for a given power. Conversely, the depth-to-width ratio increases with increasing traverse speed, though not as dramatically (Figure 10.2), and then levels off. For a given speed, the penetration depth increases with increasing power input. Generally, results obtained from low-power laser welding, say less than 5 kW, cannot be extrapolated to high-power welding, say 20 kW. This is due to differences caused by plasma shielding effects resulting from either high power densities or high evaporation rates that cause an enhanced plasma formation.

10.1.2 Effect of Beam Characteristics

The principal beam characteristics that affect the laser welding process include the beam mode, stability, polarization, and form (pulsed or CW). The effect that these parameters have on the welding process are discussed in that order in the following subsections.

10.1.2.1 Beam Mode The beam mode (Section 8.1.3) has a significant impact on the weld bead shape. A TEM_{00} mode results in a greater depth-to-width ratio than higher-order modes. This is due primarily to the more concentrated shape of the TEM_{00} mode, which results in a smaller focused

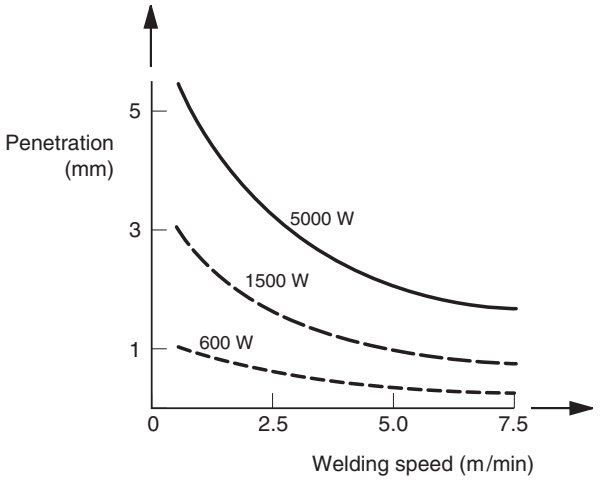


Figure 10.1 Effect of welding speed and laser power on weld penetration.

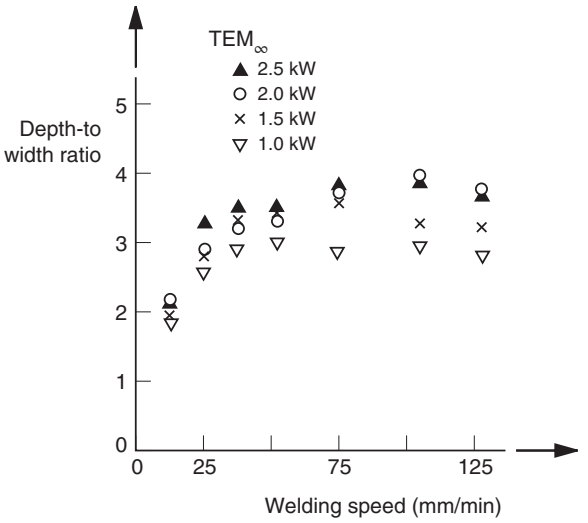


Figure 10.2 Effect of welding speed on depth-to-width ratio.

beam radius. The larger spot size of multimode beams, however, makes them more suited for butt welding applications, since it reduces the tight fit-up requirements.

10.1.2.2 Beam Stability Repeatability or consistency of the weld quality is influenced by beam stability. Temporal variations of the beam characteristics such as the output power and mode structure, as well as variations in the absorption coefficient with temperature, do cause variations in the welding process, and thus the resulting weld quality.

10.1.2.3 Beam Polarization The basic concepts of beam polarization are discussed in Section 3.4. The influence of beam polarization on the welding process depends on both the material and welding

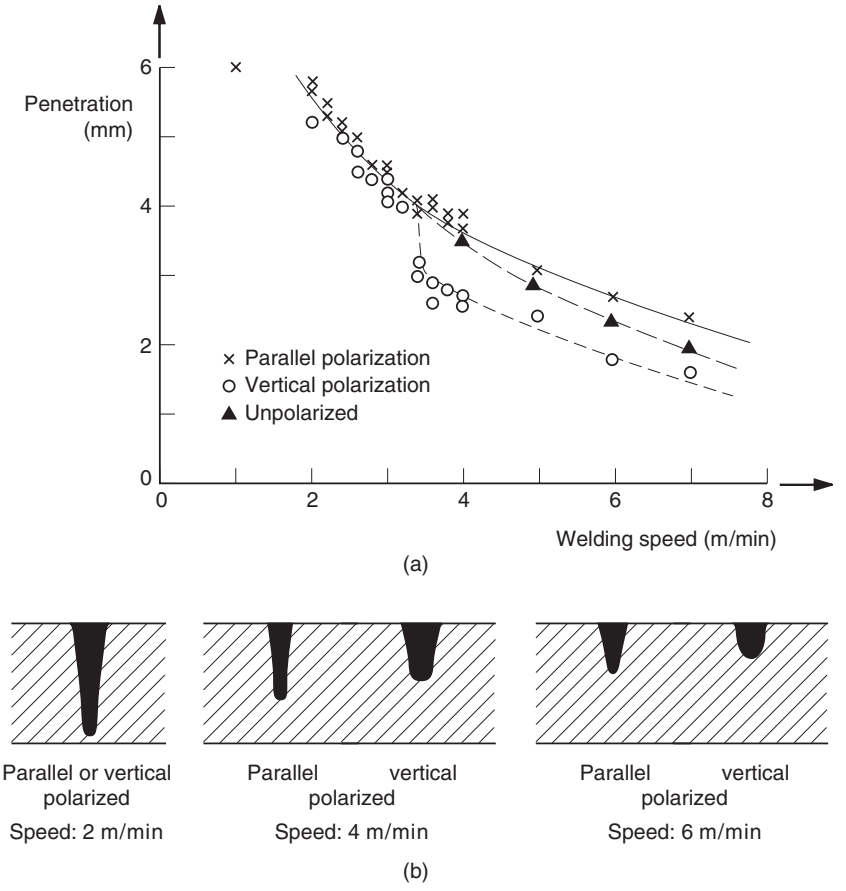


Figure 10.3 Influence of beam polarization on penetration. (a) Variation of penetration with speed for different modes of polarization. (b) Weld cross sections for parallel and vertically polarized beams for different speeds. Material – 6 mm thick steel; laser power – 5 kW; shielding gas– helium; lens f number – f6; focal diameter – 280 μm. Source: Adapted from Beyer et al. (1988).

conditions. In welding some steels, the depth of penetration is found to vary depending on the direction of beam polarization, being enhanced when the direction of polarization is aligned with the welding direction. This effect is only evident above a critical welding speed, below which the polarization direction has no impact on the penetration (Figure 10.3).

This is different from the case of laser cutting where the process efficiency is found to be strongly influenced by the beam polarization. This difference in behavior is most likely the result of a loss in correlation between any polarization of the beam and the electromagnetic field in the keyhole due to the processes of scattering, reradiation, and reflection from the walls during keyhole welding (Section 10.3.2).

10.1.2.4 Pulsed Beams Both pulsed and CW laser beams (Section 8.1.4) can be used for welding, with pulsed beams being more suited for spot welding, while CW beams are more suited for continuous welds. Pulsed beams can also be used for continuous welds if the pulses are made to overlap. Additional process parameters that need to be considered when welding with pulsed beams are the pulse duration, shape, and pulse frequency (repetition rate).

In using a pulsed beam for keyhole welding, the pulses ought to have a high enough energy to exert a significant thrust on the molten metal to maintain a keyhole once it is formed and will have to be spaced closely enough to prevent freezing of the metal while the beam is off.

When large amounts of energy are supplied in short periods to a workpiece, they tend to concentrate near the surface as heat, since it takes time for the heat to be conducted into the bulk of the material. This results in high-temperature gradients in the immediate vicinity of the heat source, producing high cooling rates once the beam is switched off.

10.1.3 Plasma Formation, Gas Shielding, and Effect of Ambient Pressure

10.1.3.1 Plasma Formation With high-power beams, a portion of the laser energy is expended in ionizing part of the shielding gas and/or vapor cloud that forms above the workpiece, to form a plasma (Figure 10.4). This plasma may absorb part of the laser beam energy and reradiate it in all directions or scatter it, thereby reducing the amount of energy reaching the molten pool. The plasma cloud may also induce an optical lensing effect that changes the effective focusing of the laser beam. This laser-induced plasma normally develops in the intensity range of 10^6 W/cm^2 while welding steel with CO_2 lasers.

Even though initial absorption of a CO_2 laser beam on a metallic surface is normally less than 10%, the absorption can be as high as 100% once a plasma is formed. Such absorption can either occur above the workpiece or inside the keyhole. When it occurs above the workpiece as described earlier, it reduces the efficiency of the welding process as a significant portion of the beam power is then lost. On the other hand, absorption inside the keyhole may enhance the process efficiency. Plasma formation is of less significance while welding with an Nd:YAG laser. This is because the Nd:YAG laser beam is more efficiently absorbed at metallic surfaces, while less is absorbed by the plasma.

The formation of the plasma has advantages and disadvantages. In the initial stages of the process, the formation of a shielding gas plasma reduces the amount of the laser energy that is reflected away as it absorbs a significant portion of the energy. Since the plasma formed is then relatively small and in close contact with the workpiece, it can enhance absorption of the beam energy by the workpiece. This, it does by first absorbing the beam energy and then transferring it to the workpiece. This is referred to as plasma-enhanced energy coupling and is most effective at the threshold or

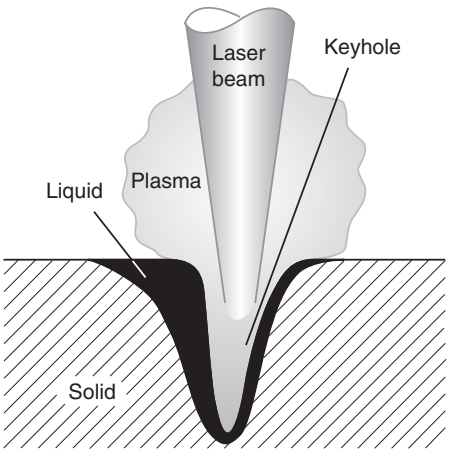


Figure 10.4 Plasma formation in keyhole welding.

initial stages of plasma formation. However, as the beam intensity increases and the plasma builds up, if the plasma stays on top of the workpiece, then the energy it absorbs is not fully coupled to the workpiece. That reduces the amount of energy that reaches or is absorbed by the work, and thus, the process efficiency.

Variations in the plasma density can thus cause variations in beam absorption by the workpiece, and as a consequence, variations in weld quality. Plasma fluctuations that occur during laser welding are more severe at low beam intensity and high speed, especially in the initial stages of plasma formation. At higher beam intensity and lower speed, the plasma is more continuous with fewer fluctuations. This is especially the case when absorption occurs inside the keyhole.

10.1.3.2 Gas Shielding In laser welding, a shielding gas may be used for one or more purposes:

1. To blow away the vapor and plasma formed, enabling the beam to reach its target.
2. To provide a protective environment for the weld pool.
3. To protect the focusing lens.

The most common gases used for laser welding are argon and helium. Argon is more commonly used for low- to medium-power laser welding since it is cheaper than helium, and its higher density results in better shielding. The argon-shielded weld surface is also smoother than that obtained with helium shielding. Since the density of helium is lower, higher flow rates are required for effective shielding. However, for high-power or low-speed applications, helium is preferable since it has a higher ionization potential. Thus, the threshold energy necessary to form a plasma when helium is used for shielding is relatively high. This reduces the amount of plasma formed and keeps the gas transparent to the laser beam, allowing the energy to reach the workpiece.

On the other hand, without the plasma, there will be no initial plasma energy coupling. This might make keyhole formation difficult, thus preventing deep penetration welding. This effect occurs, for example, in aluminum welding with a 2.5 kW laser at a speed of about 3 m/min. Under these conditions, a deep penetration weld can be obtained using nitrogen as a shielding gas, while very little molten metal is formed when helium shielding gas is used.

When argon is used in the higher-power applications, its ionization results in a significant portion of the beam being absorbed. Thus, less power reaches the weld, thereby reducing penetration. Generally, for a given set of conditions, penetration is found to be greater for helium shielding, followed by nitrogen and then argon. Nitrogen is not as extensively used as argon and helium, since it could form nitrides that might embrittle the weldment.

The weld quality thus depends on the type of shielding gas used. The effect is more significant at low welding speeds and at higher power densities, since the local vapor density above the workpiece is then much higher. As is evident from Figure 10.5, at low power levels, there is no difference between the impact of argon, helium, or nitrogen. However, at higher power levels, a distinct difference exists between the depths of penetration obtained using the various gases. As mentioned earlier, because of the high ionization potential of helium, it does not form a plasma easily. It is thus transparent (does not absorb the beam energy) to the incident laser beam over a wide range of conditions, especially at the lower welding speeds or high-power densities. By comparison, other gases tend to form a plasma that absorbs the incident beam energy, under these conditions.

Another use of the shielding gas is protection of the focusing lens. For low-power applications, say up to 1500 W, the normal flow of shielding gas (about 10 L/min) for protecting the weld is also adequate to protect the lens, as long as the gas stream is coaxial with the laser beam. A typical nozzle diameter would be 3 mm. However, at higher powers, separate shielding or helium cross-flow that is provided by a plasma suppression device (Figure 10.6) is often applied. The plasma suppression

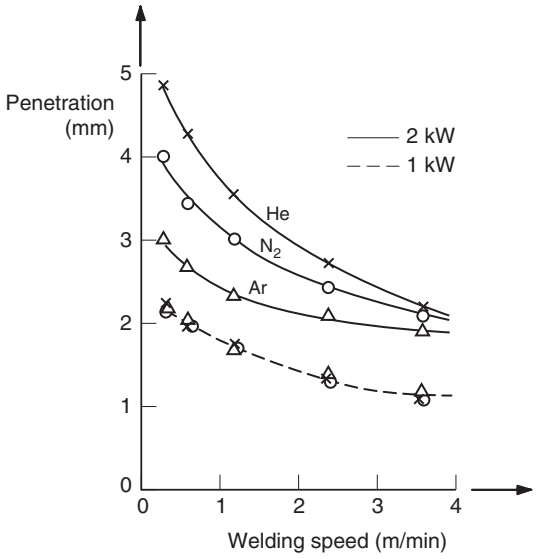


Figure 10.5 Effect of shielding gas, laser power, intensity, and welding speed on depth of penetration. Material – steel; shielding gas flow rate – 20 L/min; focused beam radius – 100 μm. Source: Adapted from Beyer et al. (1988).

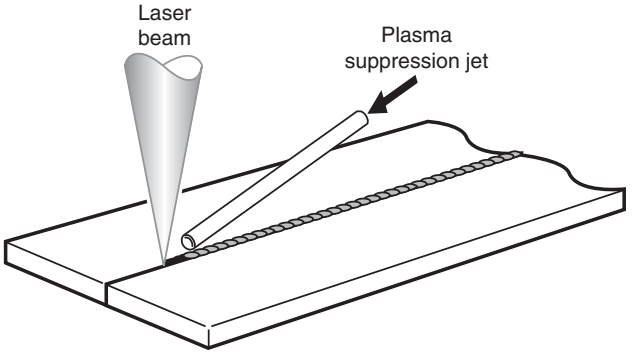


Figure 10.6 Schematic of a plasma suppression setup.

unit directs a high velocity jet of inert gas, usually helium, at the top of the weld to suppress or blow away the plasma. Plasma suppression serves to protect the lens and also to enable the heat supplied to reach the weld.

The final form of the weld bead is affected by the mode of application of the shielding gas. For example, when welding steel with a 10 kW laser while using an axial gas nozzle of diameter 2 mm, no keyhole is produced at low helium flow rates of say, less than 5 L/min, since most of the beam energy is then absorbed by the plasma. A small molten zone may result from conduction mode welding. With increasing gas flow rates, a point is reached when a transition occurs to keyhole formation, with a resulting deep penetration. Further increases in gas flow rate then produces a linear but slight increase in penetration. Beyond about 40 L/min, the gas flow may be vigorous enough to eject the molten metal out of the weld pool, resulting in an irregular and perhaps shallow bead shape. The general effect of suppressant gas flow rate on penetration is shown in Figure 10.7.

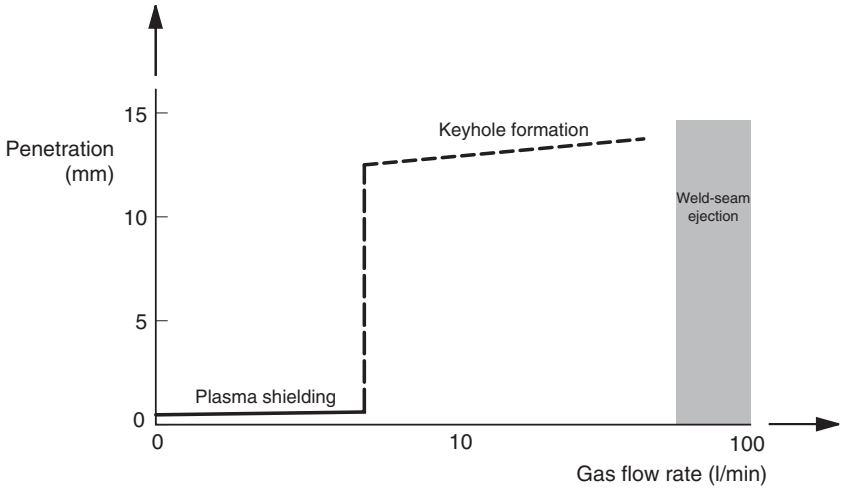


Figure 10.7 Effect of plasma suppression helium gas flow rate on depth of penetration in plain carbon steel. Laser power – 10 kW; shielding gas – helium; welding speed – 1.2 m/min; lens focal length – 476 mm. Source: Adapted from Behler et al. (1988).

10.1.3.3 Effect of Ambient Pressure The amount or intensity of plasma generated decreases as the ambient welding pressure is reduced. Below a pressure of about 5 Torr (depending on the welding conditions), the plasma is so small in size that no significant change is observed with further reductions in the pressure. The pool width decreases, while the penetration increases with decreasing pressure. Laser welding in vacuum may increase the penetration about 30–50% more than that obtained when welding in the atmosphere. However, laser vacuum welding is not a common practice due to the cost and other difficulties associated with setting up the vacuum welding system, especially for large workpieces.

10.1.4 Beam Size and Focal Point Location

For a given optical system (focusing lens), the location of the focal point relative to the workpiece surface determines the beam size on the workpiece surface, and this affects the bead width and depth of penetration (Figure 10.8). Optimum location of the focus results in maximum depth-to-width ratio, that is, the maximum penetration depth and minimum bead width occur for almost the same focal position.

10.1.5 Joint Configuration

There are various types of joint configuration that can be used in welding. However, due to the small beam size, lasers are primarily used for lap or square butt joints (Figure 10.9). The rule of thumb is that the butt gap has to be less than 10% of the thinner plate's thickness (Figure 10.10). Otherwise, there is the likelihood that most or the entire beam will pass directly through the gap, resulting in inadequate weld or no weld at all.

To minimize the problem associated with the small beam size, the beam may be defocused at the joint to produce a relatively wide bead. However, the defocused beam is more sensitive to absorption by the material surface, since the power density is then low, and any small amount of power that is reflected is relatively significant. This may result in fluctuations in the energy absorbed by

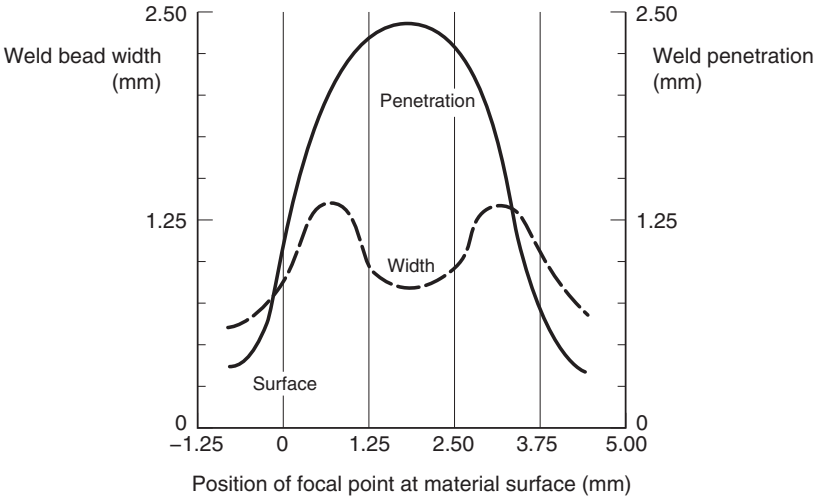


Figure 10.8 Effect of focal point positioning on penetration and bead width. Material – 1018 steel; power – 1.45 kW; welding speed – 1.25 m/min. Positive position is downward, into the workpiece. Negative is above the workpiece.

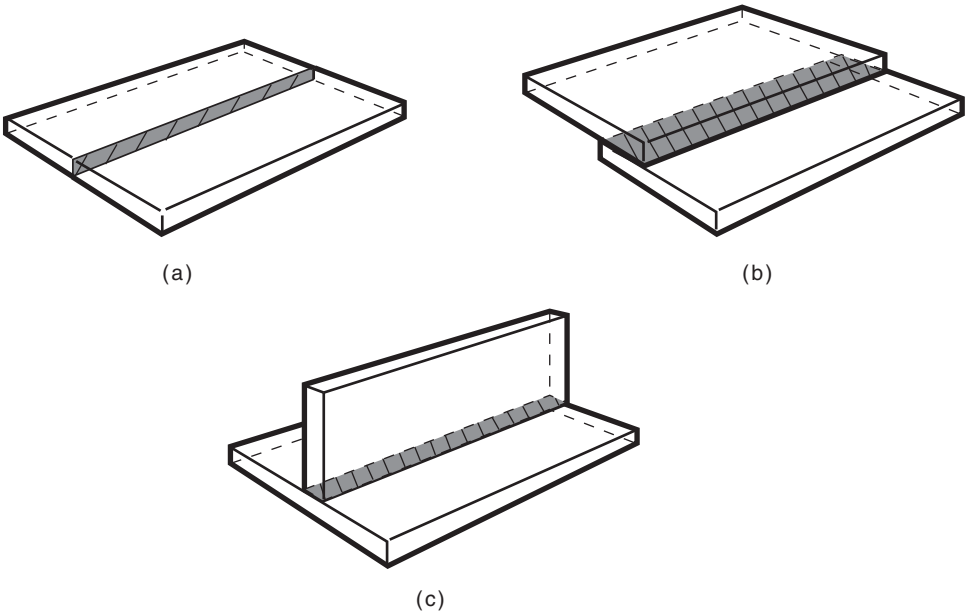


Figure 10.9 Some common joint designs. (a) Butt joint. (b) Lap joint. (c) T-joint.

the workpiece. The use of Nd:YAG lasers for metals minimizes this fluctuation, since the 1.06 μm wavelength is more efficiently absorbed by metals. Fluctuation in the energy absorbed may also be minimized by superimposing a pulsed beam over the CW welding beam. This is referred to as the rippled mode laser output or hyperpulse (Section 8.1.4). The pulse rapidly melts the material surface, thereby increasing its absorption coefficient to a high enough value that the CW beam is efficiently and uniformly absorbed.

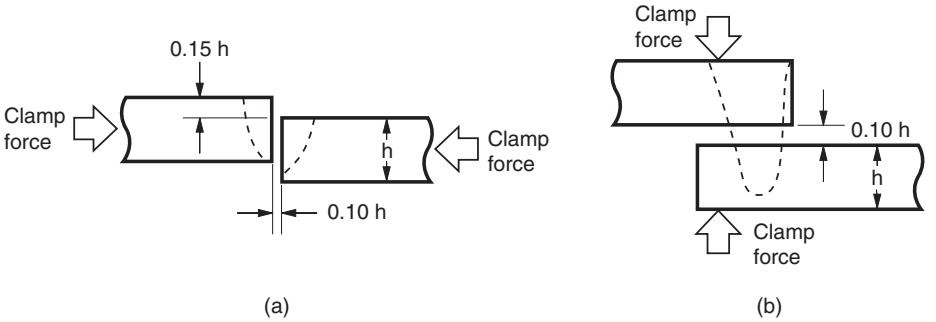


Figure 10.10 (a) Butt joint and (b) lap joint specifications in laser welding.

The joint fitting requirements for lap joints are not as critical as for butt joints since in the lap joint, the laser beam is made to penetrate through one piece into the underlying component (Figure 10.10). Lap joints are more appropriately used when the upper component is relatively thin (typically less than 1 mm in thickness). When the laser beam penetrates the top layer to join the two pieces at the interface as shown in Figure 10.10b, the process is then sometimes referred to as direct laser welding.

Thus far, a significant number of laser welds that are made are autogenous, involving no filler metal. However, filler metal may be used where necessary, especially for non-square butt joints, or when the butt gap is relatively large.

10.2 WELDING EFFICIENCY

Processing efficiency is discussed in general terms in Section 8.2. For laser welding, the overall efficiency, η, may be expressed as:

$$\eta = \frac{\text{power used in melting solid}}{\text{power input to the laser generator}} \tag{10.1}$$

Comparison of overall efficiency for arc, laser, and electron beam welding is made in Table 10.3.

10.3 MECHANISM OF LASER WELDING

When a laser beam is incident on a metallic surface, a sequence of events take place. A significant portion of the beam may initially be reflected away. The small percentage that is absorbed heats up the surface, raising its temperature. With increasing temperature, the surface absorptivity increases

Table 10.3 Comparison of efficiency values for various welding processes.

Weld type	Generator η_g %	Transmission/Heat transfer η_{th} %	Processing η_p %	Overall η %
Laser	7	70	48	2.4
Arc (argon shield)	80	50	20	8
Electron beam	80	80	48	31

Source: From Abilsiitov and Velikhov (1984)/with permission of Elsevier.

(Section 8.1.6), which further increases the temperature. This may eventually result in localized melting and possible evaporation of the metal. Such vaporization, if it occurs, creates a vapor cavity in the metal. Laser welding may thus be one of two forms:

1. Conduction mode (also known as conduction-limited).
2. Keyhole (or deep penetration) welding.

Conduction mode welding is normally used for welding of foils and thin sheets, while keyhole mode welding is used for relatively thick sections.

10.3.1 Conduction Mode Welding

Conduction mode welding usually occurs at power densities lower than 10^6 W/cm^2, with vaporization of the workpiece being minimal. The laser power is first deposited on the surface and then transferred by conduction to the surroundings. The penetration is thus essentially controlled by conduction from the initial point of contact, that is, the surface of the workpiece, causing a small surface area to be heated above the melting point. Convection also plays a role once a weld pool is formed. The resulting weld is shallower with a wider heat-affected zone compared to that produced by keyhole welding. It is more common with low-power lasers, say below 1 kW. A typical configuration is shown in Figure 10.11. The shape of the weld pool in conduction mode welding is influenced by flow in the weld pool and by the presence of surface-active elements (Section 4.3.3).

Approximate analysis of heat flow in conduction mode welding can be done using the point heat source analysis discussed in Section 4.2, while material in Section 4.3 will enable flow in the molten pool to be analyzed.

10.3.2 Keyhole Welding

At very high-power densities, say above 10^6 W/cm^2, part of the work material is vaporized to form a cavity, the *keyhole*, which is surrounded by molten metal, which in turn is surrounded by the solid material (Figure 10.12). The molten material around the keyhole fills the cavity as the beam is traversed along the joint. The cavity contains vapor, plasma, or both. Forces at play that tend to collapse the keyhole cavity are as follows:

1. Surface tension at the interface between the molten metal and vapor or plasma.
2. Hydrostatic pressure of the molten metal.
3. Hydrodynamic pressure of the molten metal.

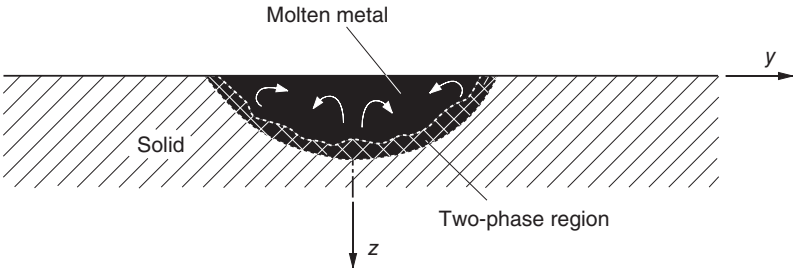


Figure 10.11 Schematic of conduction mode welding.

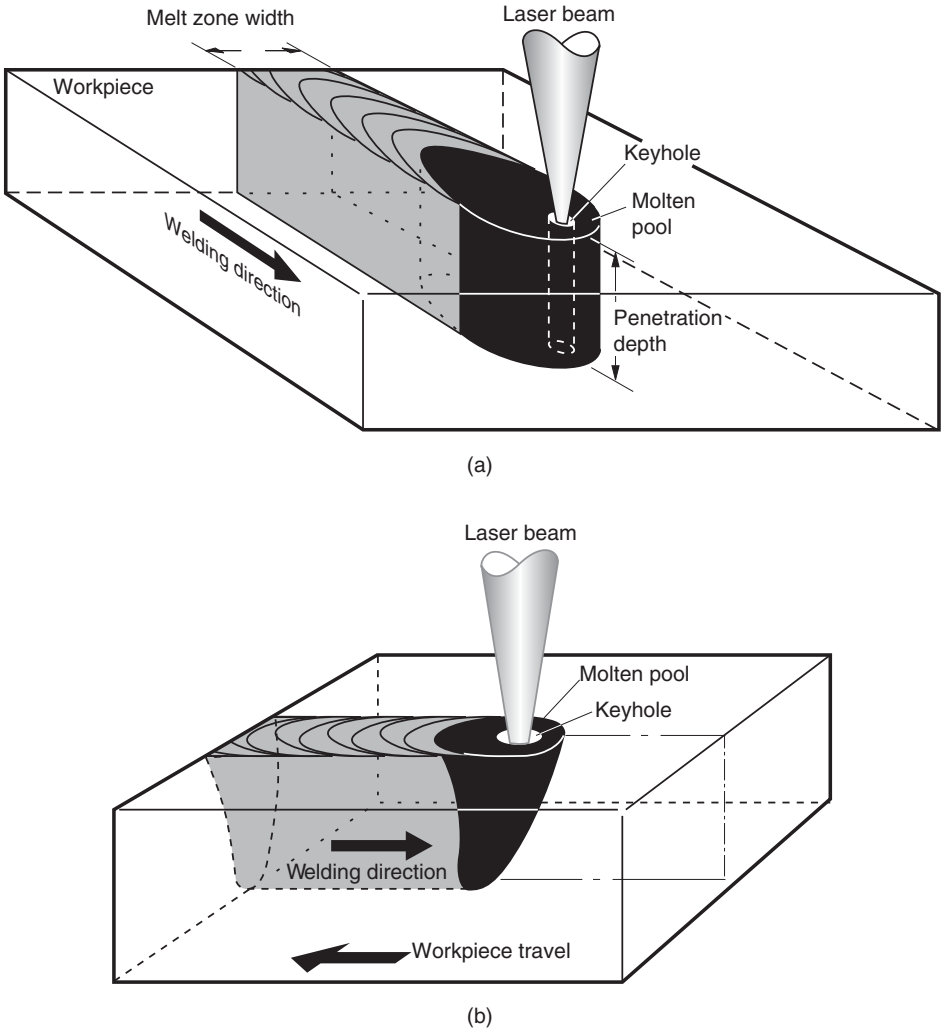


Figure 10.12 Schematic of the keyhole mode laser welding process. (a) Simplified sketch. (b) Illustration of bending of the keyhole in the direction of workpiece travel. Source: From Wilson and Hawkes (1987)/Pearson.

These forces, however, are balanced by:

1. Vapor pressure that exists within the cavity which is greater than atmospheric pressure by about 10% and which pushes the molten metal toward the periphery.
2. Ablation pressure of material as it vaporizes from the inner surface of the keyhole.

Ablation is the sudden removal of material by melting and/or vaporization. The balancing pressure is maintained during the welding process, thus sustaining the keyhole.

Laser energy that enters the keyhole is trapped and carried deeper into the material, resulting in very high absorption, greater than 90% once it is established. The keyhole mode thus enables very narrow and deep penetration welds to be made. This feature makes it attractive for machined components and aerospace engine components, since distortion is then minimal.

If the laser beam is stationary, the continuous vaporization of material would cause the cavity to grow. Under such circumstances, a steady state cannot be achieved, and the keyhole would eventually collapse. On the other hand, steady-state conditions can be achieved with a moving beam. The cavity moves along with the beam, at a rate determined by the beam. Since an adequate amount of vapor pressure needs to be maintained to prevent collapse of the cavity, a minimum traverse speed is necessary for steady-state conditions. As the cavity moves, material continuously moves from the front to the back of the cavity. This may occur either by flow of molten material around the cavity, or as vapor through the cavity or keyhole, with the bulk of the transport being through the molten material. The driving force for flow through the molten material is the ablation or recoil pressure of the evaporating material, or the surface tension due to temperature variation along the keyhole walls. The effect of surface tension as a driving force of flow is relatively small compared to the flow due to recoil pressure, since the interface temperature is close to the equilibrium evaporation temperature. The pressure in the melt at the vapor–liquid interface is a maximum directly in front of the laser beam, in the direction of welding, and reduces to a minimum on the side of the keyhole.

Under steady-state conditions, motion of the laser beam results in motion of the cavity and surrounding molten metal, along with all isotherms, at the same rate. The width of the molten region determines the width of the weld seam.

The simplest approximation to the keyhole configuration is that of a vertical cylinder (Figure 10.12a). However, since the vapor pressure generally varies with depth, the radius is also a function of depth. The variation of pressure with depth results in outward flow of vapor, causing vapor to emerge from the end of the keyhole as an axial jet. Generally, the axis of the keyhole is bent in the direction of workpiece travel (Figure 10.12b).

One typical shape of the keyhole is shown in Figure 10.13 and has a constriction near the cavity outlet. Other shapes can also develop, depending on the welding conditions. Some are relatively short with walls that are convex outward and rounded at the bottom, while others are long and thin, concave outward, and have sharp points at the bottom when blind. The shape shown in Figure 10.13 tends to occur at relatively high welding speeds and larger penetration depths (or beam powers).

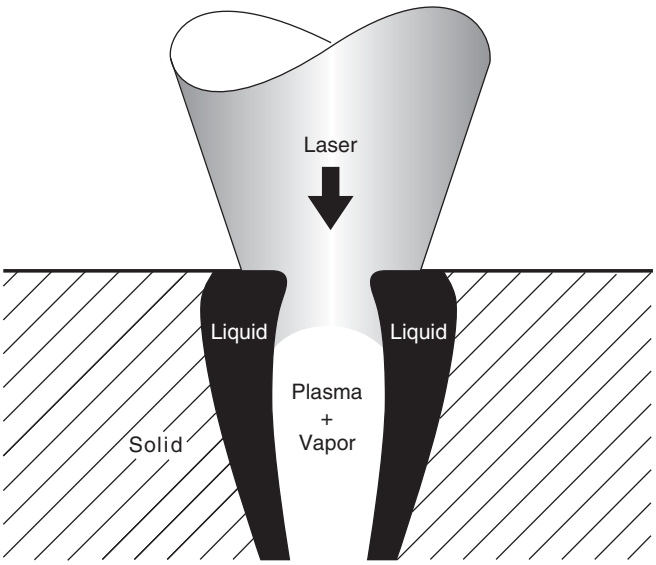


Figure 10.13 Schematic of the keyhole cavity shape. Source: Adapted from Klemens (1976)/American Institute of Physics.

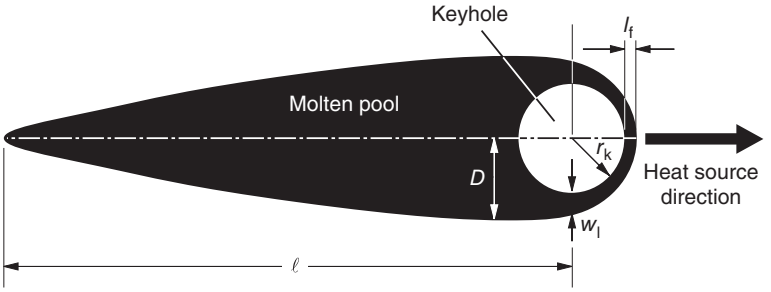


Figure 10.14 Typical top view of the weld pool during keyhole welding. Source: Adapted from Klemens (1976)/American Institute of Physics.

Due to the motion of the heat source, the top surface of the molten metal develops a teardrop shape (Figure 10.14), with a typical length-to-width ratio of about 10, and the width tails off toward the back. At low welding speeds, the molten pool is almost circular in cross section, but the teardrop shape becomes more pronounced as the speed increases (see also Figure 4.6). The width of the molten pool decreases at higher welding speeds, forcing the molten metal to move at a higher velocity through the narrow channel between the vapor and the solid boundary.

When the power level is low enough, the vapor pressure produced in the keyhole is much smaller than the fluid dynamic forces of the molten metal surrounding the keyhole. The keyhole thus collapses, and a sudden drop in penetration occurs, with the weld pool becoming roughly hemispherical in shape, resulting in conduction mode welding. Conduction mode welding is less efficient than keyhole welding, since the absence of a keyhole results in a much lower energy transfer efficiency of about 20%, as a significant portion of the incident beam is reflected away. Conduction mode welding typically results in a depth-to-width ratio of about 3 to 1 compared to about 100 to 1 for keyhole welding.

The overall transverse cross-sectional shape of the weld bead produced in keyhole welding often consists of two main sections. The lower part is nearly parallel-sided, with a gradual narrowing from top to bottom. On top of that is a region of roughly semi-circular shape, whose width is greater than that of the parallel-sided section.

10.3.2.1 Power Absorption in the Keyhole Absorption of beam energy in the keyhole occurs by two principal mechanisms:

1. Inverse bremsstrahlung of the plasma.
2. Fresnel absorption (direct deposition) at the cavity walls.

Both forms of absorption normally occur simultaneously within the keyhole and are illustrated in Figure 10.15. The absorbed beam becomes entrapped in the keyhole. This results in significantly higher absorption in deep penetration welding compared to conduction mode welding.

Inverse Bremsstrahlung Absorption With inverse bremsstrahlung absorption, a portion of the vaporized metal or shielding gas is ionized by the high incident energy to form a plasma which may absorb part of the incident beam by the inverse bremsstrahlung effect inside the keyhole. Bremsstrahlung is the electromagnetic radiation produced by the sudden retardation of a charged particle in an intense electric field. Thus in simple terms, inverse bremsstrahlung is the absorption of electromagnetic radiation as a result of the sudden acceleration of a charged particle, or absorption

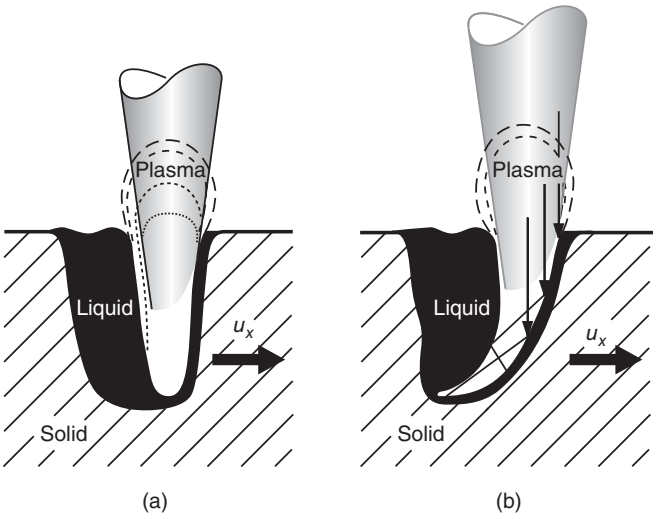


Figure 10.15 Absorption mechanisms inside the keyhole. (a) Inverse bresmsstrahlung absorption. (b) Fresnel absorption.

of a photon by an electron in the field of a nucleus. The absorption thus depends on the presence of free electrons and occurs within the plasma, which consists of ionized particles (Figure 10.15a). The absorption depends on temperature, partly because the electron density varies with temperature. As the vapor is heated in the presence of an intense beam, the absorption coefficient increases. Since the center of the keyhole is much hotter than the walls (which are at the vaporization temperature), most of the beam energy is absorbed within the cavity and then coupled to the keyhole wall through conduction and radiation.

Fresnel Absorption The mechanism of Fresnel absorption results in the beam being spread along the entire keyhole length. It results from multiple reflections at the keyhole walls (Figure 10.15b), where the beam energy that enters the vapor cavity is reflected repetitively at the walls of the keyhole until all the energy is eventually absorbed. For most applications, specular reflection (mirror-like) is assumed to be predominant, with diffuse reflection (similar to reflection from a rough surface) being neglected. Such multiple reflections, along with scattering processes, tend to randomize phase relations that may have existed in the original beam.

10.3.2.2 Keyhole Characteristics Analysis of the keyhole is quite complex and is still the subject of research. The approach we shall adopt is to first present the governing equations that are needed for an in-depth analysis. We shall then follow it with a simplified analysis which will provide an approximate insight into the keyhole characteristics.

Governing Equations In this section, we outline the governing equations necessary for analyzing the keyhole welding process, focusing on flow in the molten material. The rarefied medium in the keyhole is considered in so far as it affects the boundary conditions.

The velocity and temperature fields in the molten material, assuming incompressible flow, are given by the following governing equations:

Mass Conservation:

$$\nabla \bullet \mathbf{u_l} = 0 \tag{10.2}$$

Momentum Equation:

$$\rho_{\mathrm{l}}\left(\frac{\partial \mathbf{u_l}}{\partial t}+(\mathbf{u_l}\cdot\nabla)\mathbf{u_l}\right)=-\nabla P_{\mathrm{l}}+\mu_{\mathrm{l}}\nabla^2\mathbf{u_l}+\mathbf{F_b} \tag{10.3}$$

Energy Equation:

$$\rho_{\mathrm{l}}c_{\mathrm{pl}}\left(\frac{\partial T_{\mathrm{l}}}{\partial t}+(\mathbf{u_l}\cdot\nabla)T_{\mathrm{l}}\right)=k_{\mathrm{l}}\nabla^2T_{\mathrm{l}} \tag{10.4}$$

where $\cdot\,\mathbf{F_b}=-\rho_{\mathrm{l}}\mathbf{g}\beta_{\mathrm{T}}(T-T_0)$ is the body force vector, incorporating the buoyancy forces, with components F_x, F_y, and F_z. The negative sign results from the fact that the buoyancy force is opposite in direction to $\mathbf{g}$ (N/m^3), $\mathbf{g}$ is the gravitational acceleration vector (m/s^2), $\mathbf{u_l}$ is the velocity vector for the liquid phase (m/s), β_{T} is the volumetric thermal coefficient of expansion (/K), and s, l, and v are the subscripts for the solid, liquid, and vapor phases, respectively. For the solid substrate, only the energy equation is required.

Boundary Conditions

At the solid–liquid interface, no-slip conditions are used, that is,

$$\mathbf{u}=0$$

At the liquid–vapor interface, the boundary conditions for the momentum Eq. (10.3) are obtained by conserving momentum in the normal and tangential directions as follows:

Normal component:

$$m_{\mathrm{l}}\mathbf{n}\cdot(\mathbf{u_l}-\mathbf{u_{iv}})+m_{\mathrm{v}}\mathbf{n}\cdot(\mathbf{u_v}-\mathbf{u_{iv}})-\mathbf{n}\cdot[\mathbf{n}\cdot(\tau_{\mathrm{l}}-\tau_{\mathrm{v}})]+\mathbf{n}\cdot\nabla_{\mathrm{lv}}\gamma_{\mathrm{lv}}-(\nabla_{\mathrm{lv}}\cdot\mathbf{n})\gamma_{\mathrm{lv}}\mathbf{n}\cdot\mathbf{n}=0 \tag{10.5}$$

Tangential component:

$$\mathbf{t}\cdot\nabla_{\mathrm{lv}}\gamma_{\mathrm{lv}}-\mathbf{t}\cdot[\mathbf{n}\cdot(\tau_{\mathrm{l}}-\tau_{\mathrm{v}})]=0 \tag{10.6}$$

Energy balance at the solid–liquid interface gives the following boundary condition for the energy Eq. (10.4):

$$k_{\mathrm{s}}\mathbf{n}\cdot\nabla T_{\mathrm{s}}-k_{\mathrm{l}}\mathbf{n}\cdot\nabla T_{\mathrm{l}}=\rho_{\mathrm{s}}L_{\mathrm{m}}\mathbf{n}\cdot\mathbf{u_{il}} \tag{10.7}$$

while energy balance at the liquid–vapor interface gives the following boundary condition:

$$k_{\mathrm{l}}\mathbf{n}\cdot\nabla T_{\mathrm{l}}-k_{\mathrm{v}}\mathbf{n}\cdot\nabla T_{\mathrm{v}}+(1-R)\mathbf{I_0}\cdot\mathbf{n}=\rho_{\mathrm{l}}L_{\mathrm{v}}\mathbf{n}\cdot(\mathbf{u_{il}}-\mathbf{u_{iv}}) \tag{10.8}$$

where $\mathbf{I_0}$ is the intensity of the incident laser beam (W/cm^2), m_l and m_v are the mass flux of liquid and vapor, respectively (kg), $\mathbf{n}$ and $\mathbf{t}$ are unit normal and tangential vectors, respectively, R is the reflectivity (−), $\mathbf{u_v}$ and $\mathbf{u_{iv}}$ are the velocities of the vapor at the liquid–vapor interface and the free surface (vaporization interface), respectively (m/s), $\mathbf{u_{il}}$ is the velocity of the solid–liquid interface (melting interface) (m/s), γ_{lv} is the surface tension at the liquid–vapor interface (J/m^2 or N/m), τ is the stress tensor at the liquid–vapor interface (Pa), and ∇_{lv} is the surface gradient at the liquid–vapor interface.

The resulting Eqs. (10.2)–(10.8) are highly nonlinear, time-dependent, and coupled and can only be solved numerically. Further simplifications may still be necessary. Thus, in the following section, we present an approximate analysis which provides a simplified insight into the keyhole characteristics.

Approximate Analysis An estimate of heat flow in keyhole welding can be done using the line heat source analysis discussed in Section 4.2 to obtain the approximate temperature distribution, cooling rates, and peak temperatures, without taking into account the fluid flow (liquid, vapor, and plasma).

The discussions in Section 4.3 and the preceding section provide a basis for analyzing the molten pool that surrounds the keyhole but do not address the gaseous plume (vapor/plasma) within the keyhole.

We add to these analyses by first looking at flow conditions in a plane parallel to the workpiece surface (Figure 10.14) and make the following assumptions:

1. All variations in the z-direction (normal to the workpiece surface) over a range comparable to the keyhole radius, r_k, are small. This assumption may not be particularly valid close to the surface.
2. Welding speeds are high enough that adiabatic heating conditions exist ahead of the weld pool. In other words, heat conducted outward in the forward direction is not lost, but available to subsequently melt the material, unless it is conducted sideways, out of the path of the molten metal, faster than the advance speed. Using the nomenclature in Figure 10.14, the average speed at which heat is conducted sideways, u_{cs}, may be approximated as:

$$u_{cs} = \frac{\kappa_s}{r_k + w_l} \tag{10.9}$$

where r_k is the keyhole radius (μm), w_l is the average sideways width of the molten pool (μm), and κ_s is the thermal diffusivity of the solid (mm^2/s).

Now if the welding speed, u_x, is greater than $\frac{\kappa_s}{r_k}$, only a small fraction of the heat conducted out ahead of the molten pool will be lost. Thus, $u_x > \frac{\kappa_s}{r_k}$ defines the "adiabatic" heating condition in laser welding.

Weld Pool Length Ahead of the Keyhole, l_f The length of the molten material between the keyhole and unmelted solid in front of it can be obtained by a simple heat balance that gives

$$(c_{vol}T_m + L_{mv})u_x = k_l \frac{T_v - T_m}{l_f}\left(1 - \frac{u_x l_f}{\kappa_l}\right) \tag{10.10}$$

where $c_{vol} = \rho c_p$ is the specific heat per unit volume, assumed equal for both the solid and liquid phases (J/kg K m^3), l_f is the weld pool length ahead of the keyhole (mm), L_{mv} is the latent heat of melting per unit volume (J/m^3), T_m and T_v are the melting and vaporization temperatures, respectively (K), and κ_l is the thermal diffusivity of the liquid (molten pool) (mm^2/s).

The factor $\left(1 - \frac{u_x l_f}{\kappa_l}\right)$ corrects for the motion of the boundaries relative to the medium. Eq. (10.10) is based on the concept that the heat conducted through the molten metal provides enough thermal energy to advance the melt front at a speed u_x. Since $k_l = c_{vol}\kappa_l$, Eq. (10.10) reduces to the form:

$$l_f = \frac{\kappa_l}{u_x}\left(\frac{T_v - T_m}{T_v + L_{mv}/c_{vol}}\right) \tag{10.11}$$

from which we find that the length of the molten material between the keyhole and unmelted solid is inversely proportional to the welding speed.

Average Side Width of the Molten Pool, w_l If we divide the material around the cavity into four quadrants (Figure 10.14) and consider the first two quadrants, then in each of those two quadrants, sufficient heat must be conducted into the molten region to melt cold material at such a rate that the melt boundary in the quadrant moves with a speed u_x. The average width of this melt is w_l, and the average temperature is $\frac{T_v + T_m}{2}$. The heat conducted across the liquid–vapor interface, of projected

area r_k (assuming unit depth), under the influence of a temperature gradient $\frac{T_v - T_m}{w_l}$, provides the energy required to heat and melt the new material. Thus, a heat balance gives

$$k_l r_k \frac{T_v - T_m}{w_l} = u_x w_l \left[L_{mv} + \frac{1}{2} c_{vol} (T_m + T_v) \right] \tag{10.12}$$

$$\Rightarrow w_l^2 = \frac{\kappa_l}{u_x} r_k \left(\frac{T_v - T_m}{(T_m + T_v)/2 + L_{mv}/c_{vol}} \right)$$

$$\Rightarrow w_l \approx (l_f r_k)^{1/2} \tag{10.13}$$

Flow of Material Around the Cavity As indicated earlier in Section 10.3.2, material is transported from the front to the back side of the cavity by flow of liquid around the cavity and/or flow of vapor across the cavity. Assuming that flow occurs in the horizontal plane, then when we consider continuity of mass, we get

$$\rho_v u_{vt} = \beta_v \rho_l u_x \tag{10.14}$$

where u_{vt} is the average velocity of vapor flow transversely across the cavity (m/s), β_v is the fraction of material transported across the cavity as vapor (−), ρ_l is the density of both liquid and solid, assumed equal (kg/m^3), and ρ_v is the vapor density at temperature T_v (kg/m^3).

The ratio of the solid or liquid density to the vapor density is typically of the order of 10^4. The excess pressure that results from flow of vaporized material from the front of the cavity is due to momentum transport per unit area and unit time, which can be expressed as:

$$P = (\beta_v \rho_l u_x) u_{vt} \tag{10.15}$$

where the expression in brackets is the mass of the material removed per unit area per unit time. It is this pressure that pushes the molten metal around the keyhole cavity. The resulting velocity u_l of the molten metal can be obtained using Bernoulli's Eq. (9.1a) and is given by:

$$P = \frac{1}{2} \rho_l u_l^2 \tag{10.16}$$

From Eqs. (10.14) to (10.16), we get

$$u_l = \beta_v u_x \left(\frac{2\rho_l}{\rho_v} \right)^{1/2} \tag{10.17}$$

Now the volume of liquid (per unit time and unit cavity depth) that flows around the cavity is $2u_l l_f$, while the volume of material swept out by the cavity is $2r_k u_x$. Thus,

$$2u_l l_f = (1 - \beta_v) \cdot 2 r_k u_x \tag{10.18}$$

which gives

$$\beta_v = \frac{1}{\left[1 + \frac{l_f}{r_k} \left(\frac{2\rho_l}{\rho_v} \right)^{1/2} \right]} \tag{10.19}$$

which is the fraction of material transported across the cavity as vapor. β_v typically ranges in value between 0.01 and 0.1. From the preceding analysis, it can be shown that $\beta_v \propto u_x$, $u_{vt} \propto u_x^2$, and $P \propto u_x^4$. The analysis breaks down at high enough welding speeds u_x, as u_{vt} then approaches the speed of sound. At that point, the pressure P becomes comparable to the ambient pressure, and β_v approaches 0.5. The assumption of a pressure which is uniform over the horizontal plane is then no longer valid. Also, horizontal and vertical flow can then no longer be treated as independent.

Molten Material Behind the Cavity In estimating the dimensions of the molten pool behind the cavity, we consider the balance of energy in this region, which consists of the following components:

1. Latent heat of melting required to expand the molten material from a lateral distance of $r_k + w_l$ to D, the maximum width of the molten pool.
2. Heat liberated by condensation of the vapor transported across the cavity.
3. Heat released as the molten metal cools from a temperature of T_v behind the cavity to T_m at the edge of the molten pool.

The resulting heat balance, neglecting heat conduction from the cavity into the molten region, as well as that conducted from the liquid into the solid, which are assumed to approximately cancel out, gives

$$r_k[\beta_v L_{vol} + c_{vol}(T_v - T_m)] = L_{mv}(D - r_k - w_l) \tag{10.20}$$

where D is the maximum width of the molten zone (mm) and L_{vol} is the latent heat of vaporization (or condensation), per unit volume of liquid (J/m^3).

The length of the molten pool can be approximated by considering the distance l that the beam advances during the time t_D that it takes for the melt to attain the width D, that is, the time it takes for heat to be conducted from the rear of the cavity over a distance D. l (Figure 10.14) is then given by:

$$l = u_x t_D = u_x \frac{D^2}{\kappa_l} \tag{10.21}$$

Thus, from Eqs. (10.20) and (10.21), we have

$$l = u_x \frac{r_k^2}{\kappa_l}\left[1 + \frac{\beta_v L_{vol} + c_{vol}(T_v - T_m)}{L_{mv}} + \frac{w_l}{r_k}\right]^2 \tag{10.22}$$

Pressure Variation in the Keyhole For the purposes of this discussion, pressure within the keyhole is considered to vary only along the length of the keyhole, that is, in the vertical direction, and is uniform in a horizontal section. As indicated earlier, the cavity is kept open by an excess pressure of vapor in the keyhole, P_v (the plasma plume), and ablation pressure, P_{ab}. The forces that tend to close it are due to surface tension, P_γ, hydrostatic pressure, P_{hs}, of the surrounding molten metal, and hydrodynamic pressure, P_{hd}. Force balance thus gives

$$P_v + P_{ab} = P_\gamma + P_{hs} + P_{hd} \tag{10.23a}$$

Assuming that the principal curvature is in the horizontal plane, that is, $\frac{d^2 r_k}{dz^2}$ is negligible compared to $1/r_k$, and neglecting hydrodynamic pressure, the gage pressure (right-hand side of Eq. (10.23a)), $P(z)$, at depth z below the liquid surface is given by:

$$P(z) = \frac{\gamma_{lv}}{r_k(z)} + \rho_l g z \tag{10.23b}$$

where $r_k(z)$ is the keyhole radius at level z (μm), γ_{lv} is the surface tension at the liquid–vapor interface (N/m), and ρ_l is the density of the molten metal (kg/m^3). This pressure in the liquid is balanced by the plume and ablation pressures.

Example 10.1 A 4 kW Nd:YAG laser is used to weld an aluminum plate at a welding speed of 6 m/min. Assume that the density of molten aluminum is $\rho_l = 2400$ kg/m^3, and that adiabatic heating

conditions exist ahead of the weld pool. Further assume keyhole mode welding with the keyhole radius, r_k, being equal to twice the focused beam radius, $w_f = 125\,\mu m$.

Determine:

(1) the weld pool length ahead of the keyhole, l_f,
(2) the average side width of the molten pool, w_l, and
(3) the length of the molten pool behind the cavity, l.

Solution:

(1) The weld pool length ahead of the keyhole, l_f, is given by Eq. (10.11):

$$l_f = \frac{\kappa_l}{u_x}\left(\frac{T_v - T_m}{T_v + L_{mv}/c_{vol}}\right)$$

Then from Appendices 4D and 4E, we have

$$\kappa = \frac{k_l}{\rho_l c_p} = \frac{247}{2400 \times 900} = 114 \times 10^{-6}\ m^2/s = 114\ mm^2/s$$

$$L_{mv} = 397 \times 2700 = 1.072 \times 10^6\ kJ/m^3$$

$$c_{vol} = \rho c_p = 2700 \times 900 = 2.43 \times 10^6\ J/m^3\ K$$

$$L_{vol} = 10.78 \times 10^3 \times 2400 = 25.9 \times 10^6\ kJ/m^3$$

Therefore,

$$l_f = \frac{114 \times 10^{-6}}{6/60}\left(\frac{2494 - 660.4}{(2494 + 273) + \frac{1.072 \times 10^9}{2.43 \times 10^6}}\right)$$

$$\Rightarrow l_f = 0.65\ mm$$

(2) The average side width of the molten pool, w_l, is given by Eq. (10.13):

$$\Rightarrow w_l \approx (l_f r_k)^{1/2} = (0.65 \times 0.250)^{1/2} = 0.403\ mm = 403\ \mu m$$

(3) The length of the molten pool behind the cavity, l is obtained from Eq. (10.22):

$$l = u_x \frac{r_k^2}{\kappa_l}\left[1 + \frac{\beta_v L_{vol} + c_{vol}(T_v - T_m)}{L_{mv}} + \frac{w_l}{r_k}\right]^2$$

but

$$\beta_v = \frac{1}{\left[1 + \frac{l_f}{r_k}\left(\frac{2\rho_l}{\rho_v}\right)^{1/2}\right]} = \frac{1}{\left[1 + \frac{0.65}{0.250}(2 \times 10^4)^{1/2}\right]} = 0.0027$$

Therefore,

$$l = \frac{6000}{60} \times \frac{0.250^2}{114} \times \left[1 + \frac{0.0027 \times 25.9 \times 10^9 + 2.43 \times 10^6 \times (2494 - 660.4)}{1.072 \times 10^9} + \frac{0.403}{0.250}\right]^2$$

$$= 2.6\ mm$$

10.4 MATERIAL CONSIDERATIONS

The suitability of CO_2 laser welding to various materials is summarized in Table 10.4.

In the following sections, we shall consider the behavior of three groups of materials when laser-welded:

(1) Steels.
(2) Nonferrous metals.
(3) Ceramic materials.

In addition, we shall also discuss welding of dissimilar materials.

10.4.1 Steels

Plain carbon steels with low carbon content are generally easily welded using a laser beam. However, steels with carbon content greater than 0.25% may require preheating to prevent the formation of brittle microstructures due to the high cooling rates that normally occur during laser welding (Section 5.1.3.3). On the other hand, the high cooling rates can be beneficial in steels that depend on the addition of alloying elements to produce fine-particle strengthening, such as the high-strength low-alloy (HSLA) steels. This is in contrast to conventional arc welding of HSLA steels which generally result in strength reduction due to the greater total heat input and consequent slower cooling rates. The presence of sulfur and phosphorus as alloying elements tend to promote hot cracking as usual (Section 5.2.2.1). Some stainless steels (Section 5.1.3.3) normally result in good welds, especially most austenitic (300 series) types, except for those containing sulfur. Because of their relatively low thermal conductivities, higher depth-to-width ratios are obtained in stainless steel welds compared to carbon steels.

Galvinized steel is extensively used in the automotive, agricultural, and construction industries for protection against corrosion. However, lap welding of this material poses a significant challenge due to the low vaporization temperature of the zinc coating. When two sheets are overlapped and clamped together without any gap between them, the vapor pressure that results from zinc vaporization tends to blow out molten metal, causing spatter. The vapor also causes porosity. A common approach to overcoming this problem is to create a gap (about 0.1–0.5 mm) between the sheets by dimpling or with the use of spacers, providing an escape path for the vapor. However, this is difficult to do for complex parts, such as automotive doors.

A recent development that is supposed to overcome this problem involves the use of a fiber laser coupled to a device that enables the beam to be distributed between two regions, an outer ring that

Table 10.4 Suitability of materials for CO_2 laser welding.

Good	Fair	Poor
Low-carbon steel	Aluminum	Galvanized steel
Stainless steel	Copper	Rimmed steel
Inconel 625	Kovar	Brass
Silicon bronze	Tool steel	Zinc
Titanium	Inconel 718	Silver
Tantalum	Medium-to-high	Gold
Zirconium	Carbon steel	

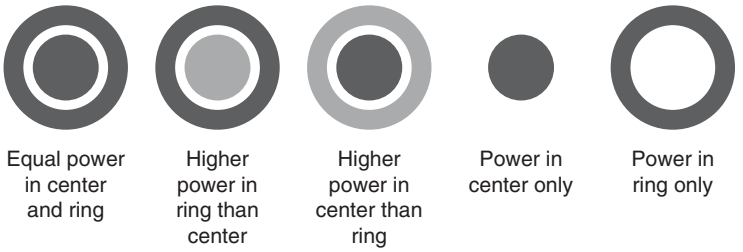

Equal power in center and ring
Higher power in ring than center
Higher power in center than ring
Power in center only
Power in ring only

Figure 10.16 A fiber laser beam distribution with an outer ring surrounding an inner Gaussian region. Source: From Kangastupa (2018)/Endeavor Business Media.

surrounds an inner Gaussian region (Figure 10.16). For the same power output, the power ratio between the two regions can be varied.

10.4.2 Nonferrous Alloys

Aluminum and copper require relatively high power densities to sustain a molten puddle due to their high thermal conductivity and reflectivity. In the case of aluminum, these properties, coupled with the low melting temperature and low ionization energy, narrow the range of process parameters that can be used for welding, compared to steel. Filler material is often necessary for welding aluminum alloys. Series 1000 and 3000 are among the easiest to weld.

Of the 2000 series, 2219 exhibits the greatest resistance to cracking because of high Cu content (6.3%). 2014 and 2024 have lower Cu (4.5%) and contain a small amount of magnesium, which increases crack sensitivity by extending the solidification range.

The 5000 series aluminum alloys with magnesium as the principal alloying element (for strengthening) tend to crack in the weld bead when welded. This is primarily due to the long solidification range which induces hot cracking (Section 5.2.2.1). The cracking problem is especially serious when the magnesium content in the weld metal is reduced to less than 3%. Silicon reduces the solidification range, thereby reducing hot cracking. Silicon is more effective when its concentration is greater than 4%. In welding the 5000 series alloys, silicon can be introduced by using an Al–Si filler material. There is also a significant loss of magnesium from the weld bead during welding of the 5000 series. This tends to reduce the mechanical properties and occurs by vaporization, since the boiling temperature of magnesium is less than half that of aluminum. Another potential problem with the 5000 series is the formation of porosity. The pores are generally spherical in nature, indicating that they are formed by gas entrapment rather than shrinkage. Since hydrogen is the primary source of such porosity, ensuring that the shielding gas is of high purity and further ensuring proper cleaning of the base metal to reduce surface contaminants and the oxide layer will tend to minimize porosity formation. The 5083 and 5456 alloys are the strongest and easiest to weld in this group.

The 4000 series with silicon as the principal alloying element are less sensitive to cracks than the 5000 series. This is mainly because the range of composition over which the alloy is crack-sensitive is narrower for the Al–Si alloys than for the Al–Mg alloys (Figure 10.17). Al–Mg–Si alloys that contain about 1% Mg_2Si are highly crack-sensitive, since the Mg_2Si content is close to the peak of the hot cracking curve (Figure 10.17). Examples are 6061 and 6063. The crack sensitivity in these alloys can be reduced by the addition of excess magnesium or silicon using Al–Mg or Al–Si filler materials, respectively. Of the 7000 series aluminum alloys, there are two main groups: (i) the Al–Zn–Mg types (for example, X7005, X7106, and 7039), which are more resistant to weld cracking,

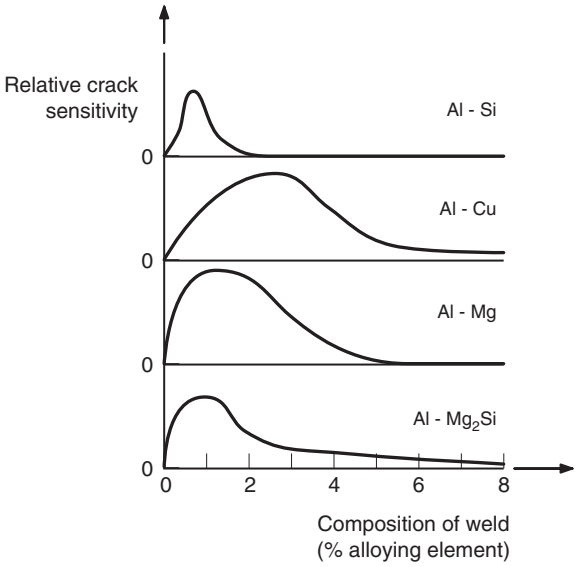


Figure 10.17 Variation of crack sensitivity with composition for various aluminum alloys. Source: From Dudas and Collins (1966)/American Welding Society.

and (ii) the Al–Zn–Mg–Cu types (e.g. 7075 and 7178), which are of higher strength, but less suitable for welding. This is because the addition of copper extends the coherence range. Crack sensitivity increases with copper content up to about 0.75% Cu and then decreases. With the Al–Zn–Mg types, the crack sensitivity increases with magnesium content up to about 1% Mg and then decreases with further increases in magnesium content.

An increasingly common application for laser welding of aluminum is in the casing of lithium ion cells used in electric vehicles, where the top is welded to the case for a hermetic seal. Tightness of the seal is important to prevent moisture from entering the cell to react with lithium and generate gas and pressure that can degrade the cell. Spatter should also be avoided during the welding process since the metal droplets can create short-circuiting within the cell.

For copper, it is often necessary to coat the surface to enhance absorption (Section 11.1.2.3). The volatile nature of the zinc in brasses makes them difficult to laser weld. Silver and gold are even more difficult to laser weld without prior treatment because of their higher reflectivity to the laser beam, especially the CO_2 laser. Alloys of nickel generally tend to absorb the CO_2 laser beam effectively, resulting in good welds, even though alloys such as Hastelloy X and Kovar are subject to hot cracking. The reactive nature of tantalum, titanium, and zirconium alloys require that they be welded in an inert atmosphere, since they are highly sensitive to oxidation.

10.4.3 Ceramic Materials

Laser welding of ceramic materials often presents problems because of the low thermal conductivity of ceramic materials and also their inherent brittleness. The low thermal conductivity, coupled with the high intensity of the laser beam, results in high-temperature gradients and thus high thermal stresses which cause the ceramic materials to crack. Welding of mullite ($3Al_2O_3{\cdot}2SiO_2$) appears to work well, provided it is preheated to a temperature of 700 °C to prevent cracking. Welding of alumina (Al_2O_3) does not present problems. On the other hand, problems may arise when silicon

nitride (Si_3N_4) is welded. Other potential problems associated with welding of ceramic materials include porosity that results from vaporization associated with heating.

10.4.4 Dissimilar Metals

Welding different materials is more difficult than welding two pieces of the same material for several reasons:

(1) When different materials are welded, differences in composition between them can result in the formation of intermetallic compounds or undesirable secondary phases if the solubility of elements of the two materials is limited. This can produce an embrittled weld joint. Phases such as eutectics and compounds that have low melting temperatures and/or are brittle can form, depending on the phase diagram of the constituent elements, and these often reduce the joint strength. For example, in laser lap welding of copper and aluminum, several different zones with diverse characteristics of morphology and composition in the intermediate layer of the weld joint can be observed (Figure 10.18). These include a banded structure with the intermetallic compound CuAl embedded in a saturated solid solution in: zone 1; zone 2 which is a combination of CuAl and $CuAl_2$ with intermetallic compounds embedded in saturated solid solution; zone 3 with rod-like and cellular $CuAl_2$ embedded in a wave eutectic; zone 4 which is a lamellar eutectic; and zone 5 which is a dendritic hypoeutectic. These different intermetallic compounds affect the mechanical properties of the joint.
(2) Differential thermal expansion resulting from different coefficients of expansion of the materials. This can create severe stresses during welding, post-weld heat treatment, or in service.
(3) Differences in thermal conductivity. This can result in improper amounts of fused metal on the two pieces. The material with the higher conductivity will have a smaller nugget size on its side of the fused metal. This can be prevented by providing more heat to the higher thermal conductivity material or preheating it.
(4) Different melting points. This may also result in the weld bead solidifying, while the lower melting base material is hot and not strong enough. This can produce hot tensile failure if the weldment is highly restrained.
(5) The formation of an electrolytic cell from the different materials. This can accelerate corrosion in service.

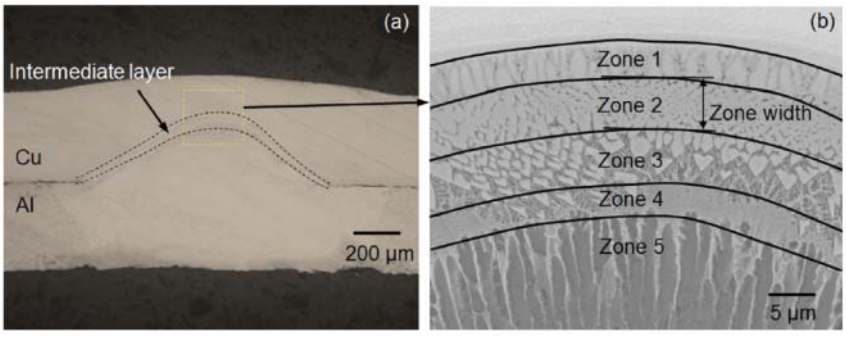


Figure 10.18 Micrograph of a Cu–Al joint cross section produced by laser lap welding with an Nd:YAG laser at a power level of 1.45–1.85 kW and welding speed 100 mm/s. (a) Optical micrograph. (b) SEM image of intermediate layer marked by the dashed square box in (a). Source: From Xue et al. (2014)/AIP Publishing.

10.5 WELDMENT DISCONTINUITIES

A more detailed discussion on discontinuities is presented in Section 5.2. In this section, we further discuss discontinuities that are commonly associated with laser welding, specifically porosity and humping.

10.5.1 Porosity

Porosity formation in high-power CW laser welding of steels depends on the welding speed (Figure 10.19). An increase in porosity with increasing speed is observed in deep penetration welding, up to a peak, and then begins to decrease at higher welding speeds. This is because in the initial stages, the increasing welding speed increases the depth-to-width ratio, which hinders the ability of the evolved gas to escape, thus increasing porosity. At much higher speeds, however, the reduced penetration enables the gas generated to escape more easily.

10.5.2 Humping

At very high welding speeds, say about 15 m/min depending on other process parameters, weld pool instability often develops, resulting in humping (Figure 10.22), and undercutting, causing irregular weld bead formation. Even though this phenomenon is not fully understood, one approach to mitigating it in electron beam welding has involved the use of a second beam to control flow within the weld pool. Whereas there has not been an analytical basis to support this methodology, extensive experimental data indicates that it is an effective means of preventing humping. The use of a second beam in laser welding has also been shown to prevent humping or increase the speed at which humping occurs. Another approach involves reducing the power input. However, the latter approach does not appear to have significant enough supporting data to warrant its widespread use. Moreover, a reduction in power input, and thus welding speed, implies a reduction in productivity.

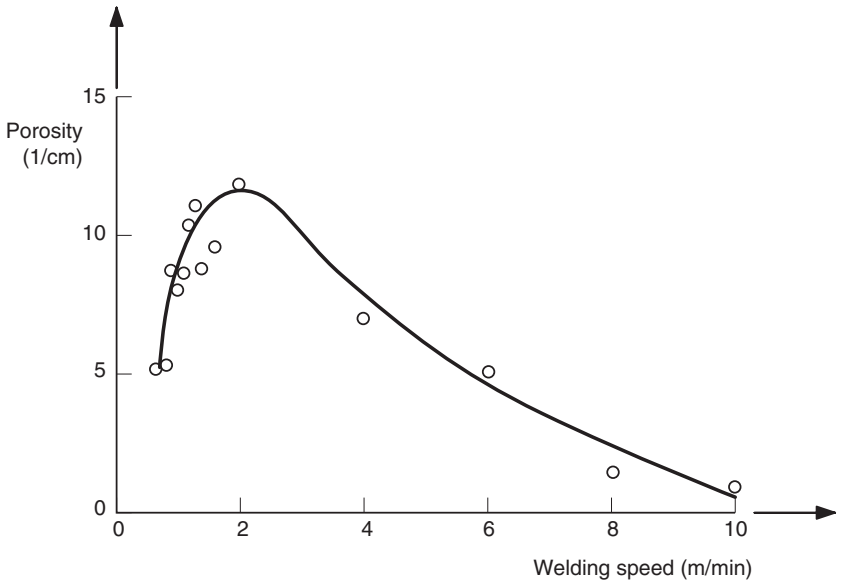


Figure 10.19 Variation of porosity formation with welding speed for a 20 mm-thick low-carbon steel welded with a 20 kW laser power. Source: From Funk et al. (1989)/SPIE.

10.6 ADVANTAGES AND DISADVANTAGES OF LASER WELDING

10.6.1 Advantages

The primary advantages of laser welding relative to conventional processes such as arc welding include the following:

1. High depth-to-width ratio, ranging from 3 : 1 to 10 : 1.
2. Low and highly localized total heat input which results in minimal distortion and small heat-affected zone size.
3. Noncontact processing.
4. The beam is neither affected by a magnetic field nor by passage through air and thus does not require a vacuum for welding as required for electron beam welding. It also does not generate radiation.
5. Easier access to the weld region, since the beam can be easily directed to ordinarily inaccessible regions. In this regard, Nd:YAG lasers are more flexible than CO_2 lasers since Nd:YAG lasers are more easily delivered to the workpiece using fiber optic cables.
6. Ability to weld dissimilar materials, since the heat input is limited to a very small region.
7. Ability to control the power density by focusing.
8. Easily automated process.
9. Relatively faster than most conventional welding processes.
10. Relatively high weld strength compared to most conventional welding processes.

10.6.2 Disadvantages

1. High cooling rates. The low energy input results in high cooling rates which may produce hardening in the weld bead and also in portions of the heat-affected zone. This is especially true of high-hardenability materials such as high carbon and some alloy steels. The formation of such undesirable microstructures is normally mitigated by either preheating before welding or post-weld heat treatment.
2. High transverse temperature gradients which may result in micro-cracking.
3. Need for precise fitting of joints. The small size of the heat source and resulting molten pool require closely and accurately fitted joints, especially in the case of butt welding since any significant opening will result in a void, undercut, or power loss through the opening.
4. High initial capital investment cost of the laser.

10.7 SPECIAL TECHNIQUES

Some of the special techniques that have been developed to enhance the quality of laser welds include multiple-beam welding, arc-augmented welding, wobble welding, and RLW.

10.7.1 Multiple-Beam Welding

Multiple-beam welding can be used either to control the microstructure in the solid material (by preheating and/or post-weld heat treating) or to control flow in the weld pool.

10.7.1.1 Multiple-Beam Preheating and Postheating As indicated in Section 10.6.2, one major disadvantage with laser welding of certain materials (for example, high carbon and some alloy steels) is that the highly intense beam results in a relatively small total heat input. As a result, very high cooling rates are achieved that can form brittle phases in such materials, and consequently, cracking. To reduce the high cooling rates, either preheating or post-weld heat treatment is used. However, conventional preheating and post-weld heat treatment have several drawbacks. In the first place, they often involve separate heating of the workpiece, usually in a furnace. This slows down the production rate, since it introduces an additional step into the process. Secondly, since the entire workpiece is often heated, energy is wasted, and also the heat applied may affect other portions of the workpiece. Finally, conventional preheating and post-weld heat treatment are often either too expensive or impractical to use, since the part may be complex in shape or bulky.

The drawbacks of traditional preheating and post-weld heat treatment can be eliminated by using the multiple-beam configuration, which, in its simplest form, is illustrated in Figure 10.20. In this configuration, there are two beams. One of the beams (the minor beam) is defocused, leads, and preheats the joint, while the major beam follows and welds the material. In another configuration, the defocused beam would follow the major beam, resulting in post-weld heat treatment. Either arrangement reduces the cooling rates, thereby reducing or preventing the formation of hard microstructures. This is especially useful in welding high hardenability materials such as high carbon and/or alloy steels and some titanium alloys.

The analysis for the dual-heat-source system parallels that for the single-source system discussed in Section 4.2. Since Eq. (4.3) is linear, we can use superposition for the two heat sources with the

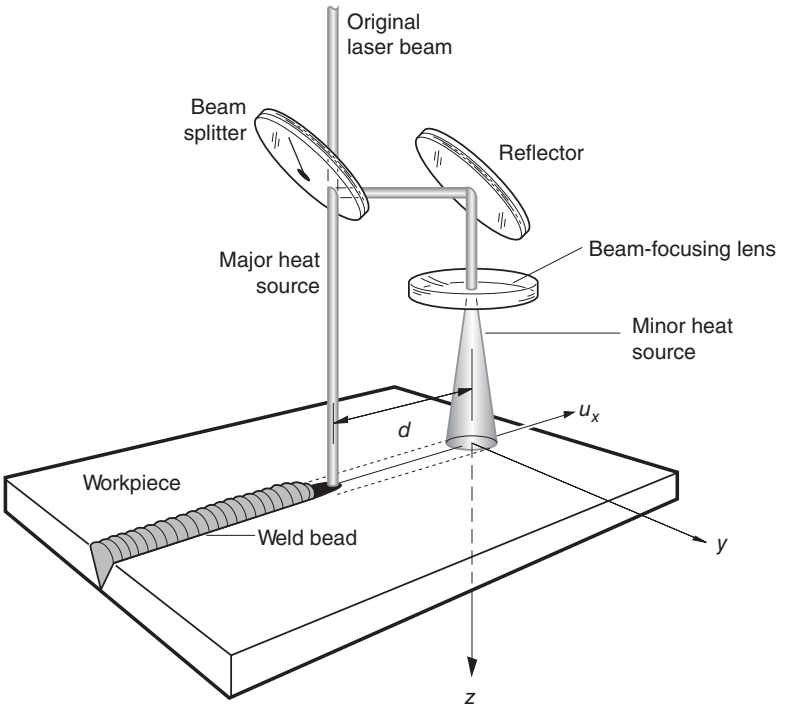


Figure 10.20 Schematic of the multiple-beam laser welding concept. Source: From Kannatey-Asibu (1991)/ American Society of Mechanical Engineers.

requirement that a common coordinate system be used. The solution for the single point source is given by Eq. (4.17) as:

$$T_1 - T_{10} = \frac{q_1}{2\pi kr} \exp\left[-\frac{u_x(r+\xi)}{2\kappa}\right] \tag{10.24}$$

This solution was obtained with the heat input at the origin of the moving coordinate system. For the minor heat source which is displaced from the origin, we obtain the corresponding solution by performing a linear transformation along the x-axis, giving

$$T_2 - T_{20} = \frac{q_2}{2\pi k\sqrt{(\xi-d)^2+y^2+z^2}} \exp\left[-\frac{u_x[(\xi-d)+\sqrt{(\xi-d)^2+y^2+z^2}]}{2\kappa}\right] \tag{10.25}$$

or

$$T_2 - T_{20} = \frac{q_2}{2\pi kr'} \exp\left[-\frac{u_x(r'+\xi')}{2\kappa}\right] \tag{10.26}$$

where q_2 is the power input for the minor heat source, $\xi' = \xi - d$, and $r'^2 = (\xi-d)^2 + y^2 + z^2 = \xi'^2 + y^2 + z^2$.

Superposing the solutions for the two heat sources from Eqs. (10.24) and (10.26), we get

$$\Theta = \frac{1}{2\pi k}\left[\frac{q_1}{r} e^{\left[-\frac{u_x(r+\xi)}{2\kappa}\right]} + \frac{q_2}{r'} e^{\left[-\frac{u_x(r'+\xi')}{2\kappa}\right]}\right] \tag{10.27}$$

where $\Theta = T_1 + T_2 - T_{10} - T_{20}$.

In analyzing the cooling rate, we again consider only the worst-case scenario, which occurs along the bead centerline, that is, $y = z = 0$. Thus from Eq. (10.27), we have

$$\Theta = \frac{1}{2\pi k}\left[\frac{q_1}{|\xi|} e^{\left[-\frac{u_x(r+|\xi|)}{2\kappa}\right]} + \frac{q_2}{|\xi'|} e^{\left[-\frac{u_x(r'+|\xi'|)}{2\kappa}\right]}\right] \tag{10.28}$$

With the moving heat source, cooling does not start until after the source passes the point of interest. In this dual-heat-source case, one would expect the cooling to start after the leading source passes the point. However, since the lead or minor source is only being used for preheating, and the major source will reheat points between the two sources, most likely to a temperature higher than that experienced as a result of the minor source, we are only interested in cooling rates along points behind the major source, that is, for $\xi < 0$. This reduces Eq. (10.28) to

$$\Theta = \frac{1}{2\pi k}\left[\frac{q_1}{\xi} + \frac{q_2}{\xi'}\right] \tag{10.29}$$

Differentiating with respect to time, we have

$$\frac{d\Theta}{dt} = \frac{1}{2\pi k}\left[q_1 \frac{d}{d\xi}\left(\frac{1}{\xi}\right)\frac{d\xi}{dt} + q_2 \frac{d}{d\xi'}\left(\frac{1}{\xi'}\right)\frac{d\xi'}{dt}\right] \tag{10.30}$$

But $\dfrac{d\xi}{dt} = \dfrac{d\xi}{dx}\dfrac{dx}{dt} = u_x$ Likewise, $\dfrac{d\xi'}{dt} = u_x$

Thus, the cooling rate becomes

$$\frac{d\Theta}{dt} = -\frac{u_x}{2\pi k}\left[\frac{q_1}{\xi^2} + \frac{q_2}{\xi'^2}\right] \tag{10.31}$$

Without loss of generality in the basic deductions to be made, we further consider the situation where the two heat sources are of the same intensity, which reduces Eq. (10.31) to the form:

$$\frac{d\Theta}{dt} = -\frac{q_1 u_x}{2\pi k}\left[\frac{1}{\xi^2} + \frac{1}{\xi'^2}\right] \tag{10.32}$$

But $(2\pi k)^2\Theta^2 = {q_1}^2\left[\frac{1}{\xi^2} + \frac{1}{\xi'^2} + \frac{2}{\xi\xi'}\right]$, from Eq. (10.29).

Therefore,

$$\frac{d\Theta}{dt} = -\frac{q_1 u_x}{2\pi k}\left[\frac{(2\pi k)^2\Theta^2}{{q_1}^2} - \frac{2}{\xi\xi'}\right] = -\left(\frac{2\pi k u_x \Theta^2}{q_1} - \frac{q_1 u_x}{\pi k \xi\xi'}\right) \tag{10.33}$$

The first term on the right-hand side of Eq. (10.33) is equivalent to the cooling rate at a temperature Θ obtained for a single heat source. As is evident from the equation, introduction of the second heat source results in a reduction term in the cooling rate equation. In other words, at a given temperature, the dual-heat-source system reduces the cooling rate by an amount equal to the second term on the right-hand side. Thus, at points in a dual-heat-source system and a single-source system which have the same temperature (not necessarily the same location), the cooling rate will be lower for the dual-source system, even if the heat input for the single-source system is the same as each of the dual-system inputs.

The defocused preheating beam is useful when welding thin sheets. For relatively thick materials, it might be necessary to have both beams focused and being used for penetration welding. The leading beam then welds and at the same time preheats the material for the follow-up beam.

The use of the multiple-beam system also results in a reduction in the steep transverse temperature gradients $\left(\frac{dT}{dy}\right)$ that arise during the welding process. This is especially useful in welding or cutting inherently brittle materials such as ceramics, where the steep temperature gradients can induce high thermal stresses and thus cause cracking of the workpiece.

The effect of the multiple-beam system on the weldment depends on the power ratio of the two beams, the inter-beam spacing, and the distribution of the two beams. As Figure 10.21 illustrates, for a given power of the major beam, increasing the power of the preheating beam reduces the cooling rate for any temperature. The transverse temperature gradient is also reduced. It can be shown that the lowest cooling rate is obtained when the preheating beam is behind the major beam, that is, the inter-beam spacing is negative. That is essentially equivalent to post-weld heat treatment and implies that in using the multiple-beam system, post-weld heat treatment would result in a lower cooling rate than can be obtained from preheating. On the other hand, the highest transverse temperature gradients are obtained when the inter-beam spacing is negative. In other words, post-weld heat treatment is not as effective for reducing the transverse temperature gradients as preheating.

Thus, the configuration that is selected for a given application should be determined by the material to be welded. For example, for alloy steels where the martensitic transformation can be a problem, post-weld heat treatment, which effectively reduces cooling rates, should be used, whereas for ceramics where the thermal stresses caused by transverse temperature gradients can present problems, preheating would be more appropriate.

The distribution of the leading beam intensity affects neither the cooling rate nor transverse temperature gradient as significantly as does the power ratio. However, the beam distribution tends to affect the transverse temperature gradient more than the cooling rate.

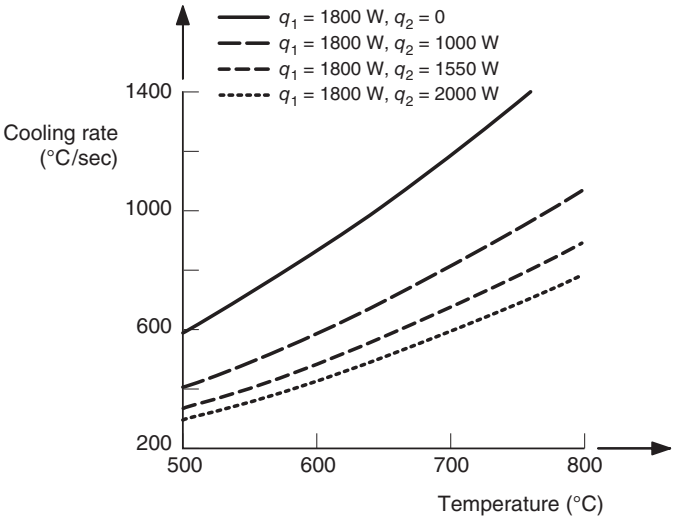


Figure 10.21 Effect of preheating with a dual beam on cooling rates. Source: Adapted from Liu et al. (1993).

Example 10.2 A CO_2 laser output is split into two 3 kW beams and used for butt welding of two semi-infinite plates of AISI 1080 steel, each 6 mm thick. A dual-beam system is used to prevent the formation of a brittle microstructure in the weld bead. Assume that 20% of each beam is lost by reflection.

If the critical cooling rate to avoid the formation of martensite in AISI 1080 is about 150 °C/s at a temperature of 550 °C, and the minor heat source has been set at a distance $d = 4$ mm from the major heat source, what is the maximum welding speed, u_x, that can be used to avoid martensite formation?

Solution:
The cooling rate for a dual-beam system in a semi-infinite plate is given by Eq. (10.32):

$$\frac{d\Theta}{dt} = -\frac{q_1 u_x}{2\pi k}\left[\frac{1}{\xi^2} + \frac{1}{\xi'^2}\right]$$

Thus to determine the welding velocity, we first have to determine the positions at which the temperature of 550 °C is reached. From Eq. (10.29), we have

$$\Theta = \frac{1}{2\pi k}\left[\frac{q_1}{\xi} + \frac{q_2}{\xi'}\right]$$

$$\Rightarrow 550 = \frac{1}{2\pi \times 73}\left[\frac{0.8 \times 3000}{|\xi|} + \frac{0.8 \times 3000}{|\xi - 0.004|}\right]$$

which gives

$$\xi = 21.2 \text{ mm}, -17.2 \text{ mm}.$$

However, since we are interested in phenomena that occur during cooling of the weld, our focus is on temperatures behind the major heat source, that is, $\xi = -17.2$ mm. The cooling rate at this

location should not be greater than 150 °C/s. Thus, we have

$$-150 = -\frac{0.8 \times 3000 \times u_x}{2\pi \times 73}\left[\frac{1}{(-0.0172)^2} + \frac{1}{(-0.0172 - 0.004)^2}\right]$$

$$\Rightarrow u_x = 0.025 \text{ m/s} = 25 \text{ mm/s}.$$

10.7.1.2 Multiple-Beam Flow Control A number of discontinuities such as humping and undercutting that arise during high-intensity welding processes are associated with the nature of molten metal flow within the weld pool. Extensive research has been done using multiple beams to minimize these defects in electron beam welding. The following discussion is based on the electron beam research results. It is expected that in general, it will also apply to laser welding, given similarities in the beam intensities.

At relatively high speeds, say greater than 15 m/min in the keyhole mode, humping can occur (Figure 10.22). The pool width decreases, and the molten metal is forced to move at higher speeds through the narrow channel between the lateral walls and the keyhole (Figure 10.23a).

In regions where the joint is not exposed, humping might appear to be esthetically acceptable or could simply be machined off. However, beneath the hump often exists an opening or lack of penetration which might result in structural weakness. It is difficult to suppress these discontinuities in single-beam welding without compromising on the welding speed and penetration. However, such defects have been shown to be effectively curtailed or suppressed by using two beams in series along the welding direction in electron beam welding. One beam, B-1 (the leading or major beam, beam 1), is used as a heat source to melt the specimen with full penetration, while the second beam, B-2 (the trailing or minor beam, beam 2), is used to control flow of the molten metal (Figure 10.23b). Using the same welding conditions that produce a humped bead at 15 m/min with a single beam, speeds of up to 30 m/min can be achieved with the multiple-beam system.

Important parameters associated with multiple-beam processing include the inter-beam spacing and power ratio. The location of B-2 with respect to the molten pool produced by B-1 affects the quality of the bead produced and can be classified into three regions, region A, region B, and region C (Figure 10.23b). In region A, the inter-beam spacing, l_b, is less than the cavity length, l_c, that is, $l_b < l_c(s)$, which means the second beam is located in the cavity formed by the leading beam. Humping and undercutting are enhanced under such circumstances. This is because the recoil pressure due to the vapor resulting from the second beam accelerates the flow rate of the molten metal in the pool cavity, thereby further elongating the cavity. In region B, l_b, is less than the distance between a single beam and its hump, $l_h(s)$, that is, $l_c(s) < l_b < l_h(s)$. In other words, the second beam impinges on the surface of the molten pool between the pool cavity and the hump that would otherwise be produced by the single beam. This is an effective way to prevent the hump formation. Under these

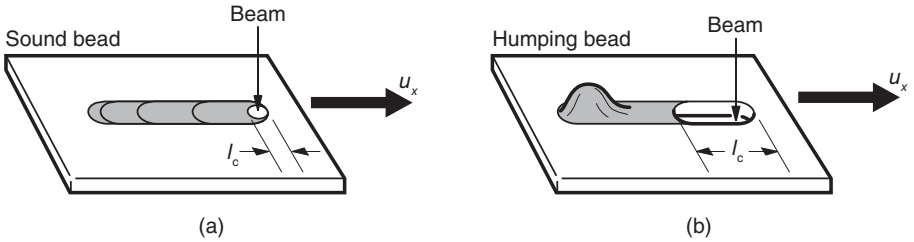


Figure 10.22 Schematic of humping during single-beam welding. (a) Sound bead. (b) Bead with a hump. Source: From Arata and Nabegata (1978)/with permission of Osaka University.

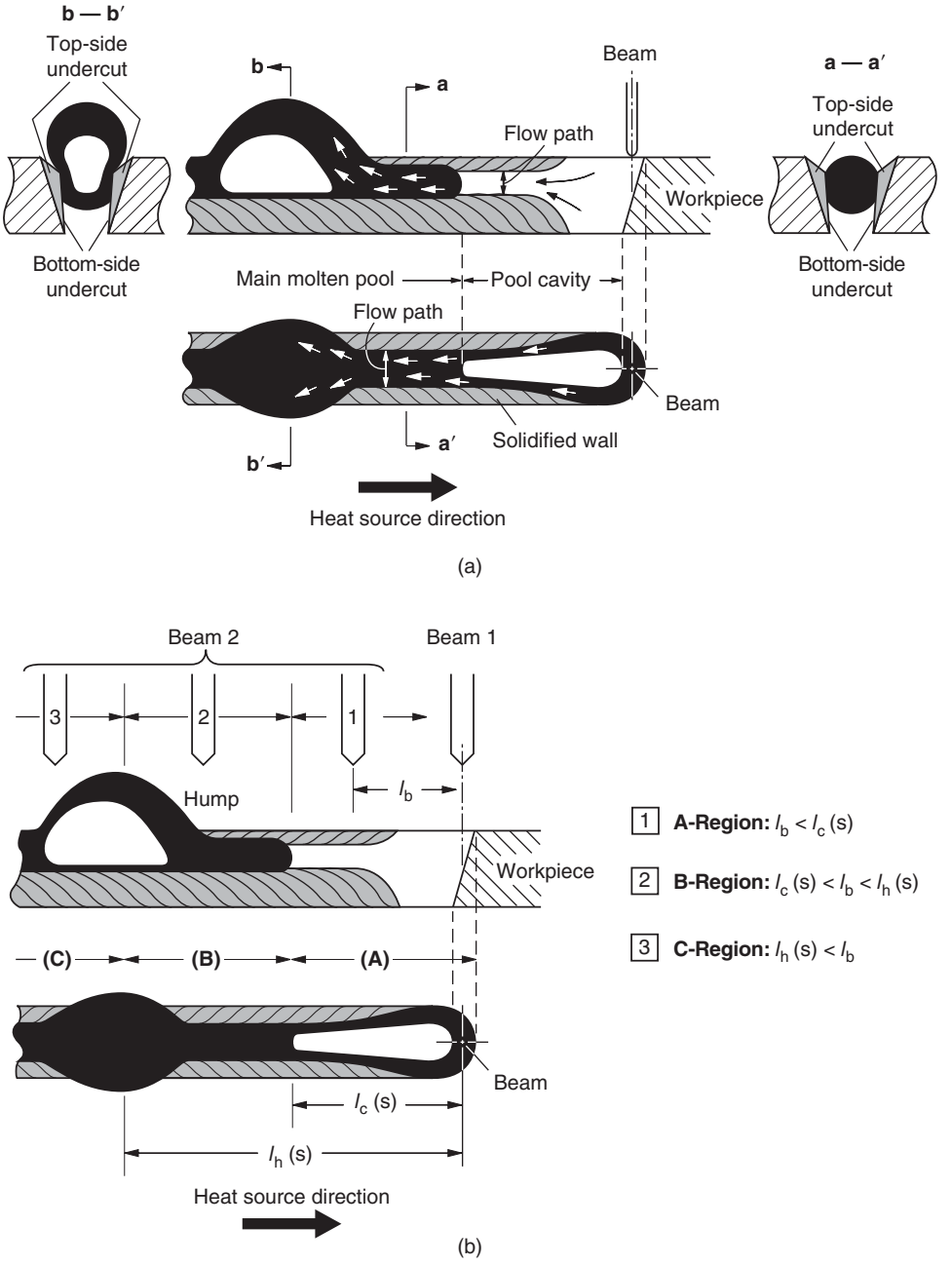


Figure 10.23 (a) Mechanism of humping during single-beam welding. (b) Prevention of humping using dual-beam welding. Source: From Arata and Nabegata (1978)/with permission of Osaka University.

circumstances, the flow direction of the molten metal is changed toward the lateral walls, resulting in melting of the solidified walls at the top and bottom of the lateral walls. This in turn broadens the flow in the molten pool and, consequently, reduces the inner pressure of the pool and prevents hump formation. With this configuration, the pool cavity length also becomes shorter. In region C,

the second beam is located much further down, beyond where the hump would start to form for the single beam, and the bead formation again becomes irregular.

In addition to the location, proper selection of the beam power ratio is also important in achieving good weld quality. If the power ratio (B-2 to B-1) is too low, broad undercutting results, and this is because the flow generated by the second beam is not adequate to fill up the top and bottom parts of the lateral walls with molten metal. On the other hand, if the power ratio is too large, an irregular weld bead with undercutting results, as the second beam then widens the pool width. The optimum power ratio of B-2 to B-1 in terms of pool uniformity (size from the top to the root) and stability has been found to be about 10%.

10.7.2 Arc-Augmented Laser Welding

Laser welding can be significantly improved if combined with arc welding. Generally, for a given total energy input to the process, the augmented process is found to produce a greater penetration than either arc welding or laser welding by itself. For example, if a 5 kW laser beam is combined with 300 amp gas tungsten arc welding (GTAW), the resulting depth of penetration is about 1.3 to 2 times that obtained using the 5 kW laser alone. Similarly, higher penetration is obtained with the laser-augmented gas metal arc welding (LAGMAW) process than with gas metal arc welding (GMAW) alone. The increase in penetration associated with LAGMAW is more significant at low arc welding currents (Figure 10.24a). The reduction in penetration at higher currents could be due to absorption of the laser beam by the increased plasma formation.

One possible setup is illustrated in Figure 10.24b. It is desirable to have the distance between the laser beam and the electrode as close as possible. In the case of GTAW, no significant increases in penetration is observed when the separation is less than 3 mm and the laser power is 2 kW with a current of 300 amps, and a welding speed of 1 m/min. At greater distances, however, the penetration reduces with increasing separation distance. In addition to increased penetration, another advantage of the augmented welding process is the elimination of undercut and humping which can occur at high speeds during GTAW alone. One disadvantage of arc-augmented welds is that they tend to have a larger heat-affected zone size than a weld made using a laser of the same total power.

The bead appearance is found to be better when the laser beam leads the arc in LAGMAW. This is because the suppression gas flow does not affect the molten pool created by the arc. However, a more stable arc is observed when the laser beam follows the arc.

10.7.3 Wobble Welding

The laser welding process can be enhanced using the wobble head design, which incorporates a scanner within the weld head. This superposes a predefined motion of the focal spot on the weld centerline motion defined by the joint trajectory, making the beam motion dynamically adjustable. The resulting process is known as laser wobble or stir welding. For a circular motion, the scanned beam speed, u_b, is given by:

$$u_b = \pi D \nu \tag{10.34}$$

where ν is the wobble frequency and D is the wobble diameter. The overall beam speed is then determined by u_b and the specified welding speed, u_x. However, u_b is generally much higher than u_x and thus significantly influences the weld dynamics.

Any geometric pattern can be used for the wobble, but common designs are illustrated in Figure 10.25a. For a constant energy input, increasing the amplitude of wobble oscillation (D)

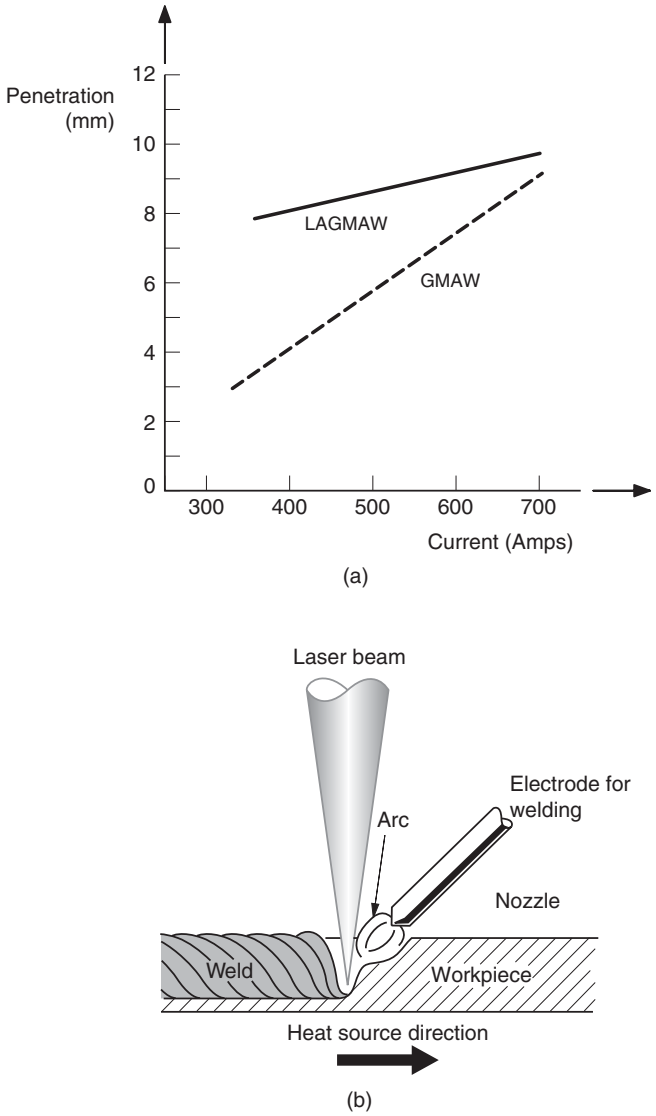
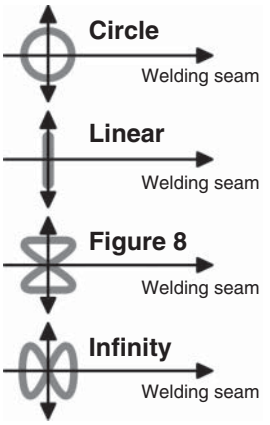


Figure 10.24 (a) Effect of welding current on laser-augmented arc welding. Source: From *ICALEO 1990 Proceedings*, Vol. 71. Copyright 1991, Laser Institute of America. All rights reserved. (b) Laser arc welding system with both laser beam and arc on same side of the workpiece. Source: From Duley (1983)/with permission of Springer Nature.

changes the weld cross-sectional geometry from a traditional nail head to a rectangular profile (Figure 10.25b), enabling the weld cross section to be effectively controlled. This may involve a transition from a keyhole mode to conduction mode welding.

Figure 10.26 compares the welds obtained for butt, lap, and fillet welds when using wobble and traditional laser welding processes. As can be observed in the figure, welds produced for fillet and lap joints by traditional welding resulted in porosity, while the corresponding welds for wobble welding were pore-free.



(a)

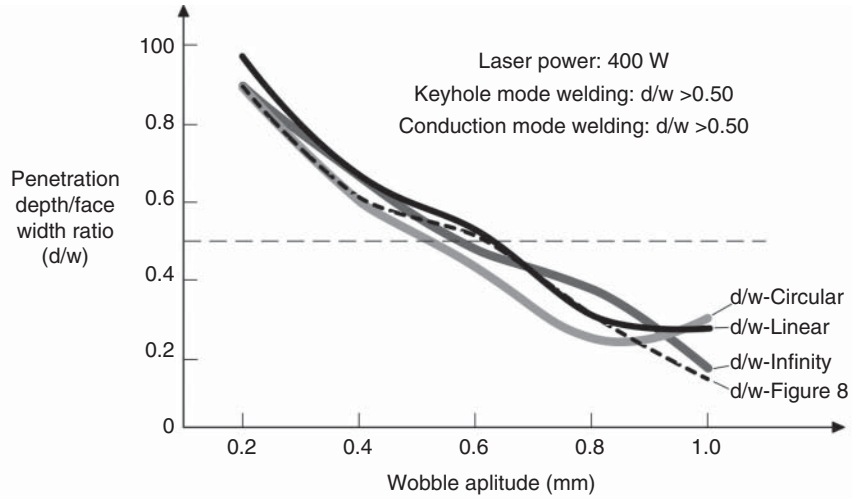


Figure 10.25 Illustration of wobble welding. (a) Common wobble patterns. (b) The effect of wobble oscillation amplitude on weld geometry. Source: From Kancharla et al. (2018)/Endeavor Business Media.

Wobble welding is more effective when the focused beam size is smaller, since a higher power density is then obtained at the focal point. This makes it possible to initiate a keyhole while welding high-reflectivity materials such as copper and aluminum using a near-infrared wavelength (about 1 μm) laser and yet be able to fill the joint with the smaller beam. In other words, joint fit-up requirements do not need to be as stringent. Smaller focused beam sizes (about 20 μm) are achievable with single- and low-mode near-infrared fiber lasers.

Optimal selection of wobble parameters minimizes the potential for forming weld discontinuities such as porosity and cracking.

10.7.4 Remote Laser Welding

RLW enables the welding process to be accomplished using long focal length optics (up to 1600 mm, Table 10.2). The basic concept is illustrated in Figure 10.27a. The optics consists of two lenses and

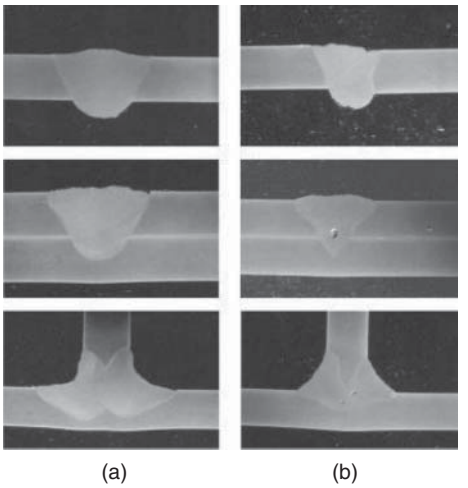


Figure 10.26 Comparison of welds produced by: (a) wobble welding and (b) traditional laser welding for butt, lap, and fillet welds on 3 mm thick 6013-T4 aluminum alloy. Source: From Richard Martukanitz and Israel Stol (2017), Endeavor Business Media, LLC.

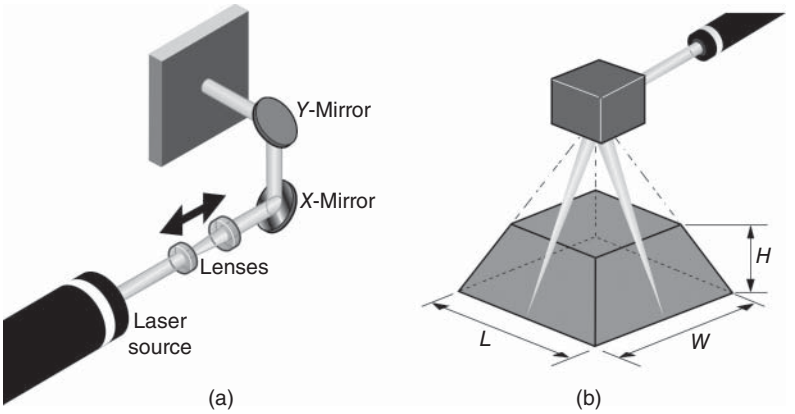


Figure 10.27 (a) Beam delivery system for remote laser welding. (b) Work envelope.

two mirrors, with the first lens being moved along the optical axis to achieve correct positioning of the focal point as the beam moves along the workpiece surface. The two mirrors move the beam in the x and y directions. There are two forms of the beam delivery system:

1. Stationary head.
2. Robotic head.

Stationary Head Positioning of the focal point within the work envelope can be achieved with both the workpiece and scan head stationary. The only components that move are the first lens and the two mirrors. This has the advantages of accuracy, lower cost, and simplicity of programming. The drawback is that the beam might not have access to certain parts of the workpiece, as the line of sight might be blocked by part of the object.

Robotic Head The scanning system in this case is mounted at the end of a robot. Thus, all locations of the workpiece surface are accessible. However, it is less accurate (since inaccuracies of the robot contribute to the overall system error), more expensive, and more complex to program. Also, the robot arm inertia slows down the system.

For both head types, the work envelope is in the form of a truncated pyramid (Figure 10.27b) and is usually defined such that the maximum inclination of the beam to the work surface normal is 20°.

RLW is especially useful in applications where joint accessibility is difficult, such as joints within a muffler. It is also a good replacement for a number of spot welding applications and is thus being increasingly used in the automotive industry. Major drawbacks of RLW are cost and quality. The high initial cost is such that it only becomes more economical than resistance spot welding when a large number of welds, say five million, are to be produced per year. Depending on the product design, the quality of some of the welds produced by RLW can vary with position because of variations in the beam orientation, and thus the area of incidence.

10.8 SPECIFIC APPLICATIONS

Two applications of laser welding that have received quite a bit of attention in recent years are microwelding and welding of tailored blanks.

10.8.1 Microwelding

The small size to which the heat source can be focused makes laser welding suitable for welding and soldering small components or joints as found in electronic components, medical devices, or in attaching a thermocouple to a part for temperature measurement. The ability to weld dissimilar metals using a laser beam and also the minimal thermal damage to surrounding material make it particularly attractive in this regard. For such applications as microwelding in electronic circuitry, the shorter wavelength fiber and Nd:YAG lasers make them preferable to the CO_2 laser (10.6 μm wavelength). The shorter wavelength beam is least absorbed by insulating materials and thus can be scanned from one solder joint to another without shutting off the beam. In contrast, the longer wavelength beam would need to be cycled on and off as the beam traverses from one joint to another as it can damage the intermediate insulating material.

Specific examples of microwelding applications include micro-spot welding of razor head blades together and hermetic instruments such as heart pace makers and transducers.

10.8.2 Laser-Welded Tailored Blanks

Tailored blank welding essentially involves welding of sheet metal blanks of either different thicknesses (gages), coatings (bare or galvanized), or materials (grades). There are two primary ways in which tailored blanks are made:

1. Resistance mash-seam welding.
2. Laser welding.

This section focuses on laser-welded tailored blanks. Tailored blanks have replaced a number of units that are traditionally produced by first forming single components, which are then subsequently joined by welding, especially for the automotive body-in-white. One typical application of tailored blanks in this area is the use of a thicker or higher strength material to replace the reinforcement that would normally be required for strength or support. An example is the center pillar inner of a car body (Figure 10.28), where a thicker material is used in the upper portion where a separate reinforcement

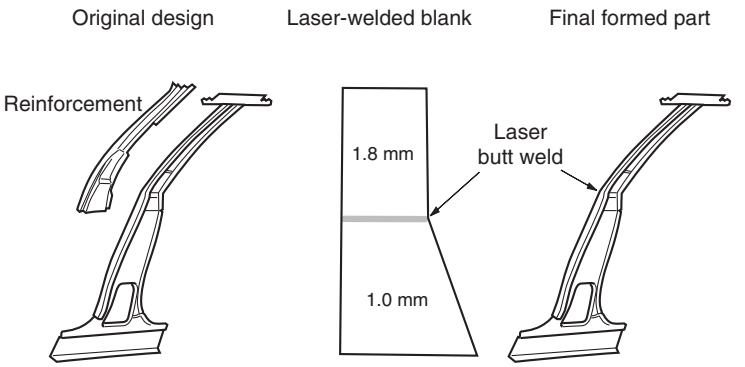


Figure 10.28 Cadillac center pillar inner. Source: From Uddin (1993)/Laurin Publishing Company.

would normally be needed and joined to the thinner material that is subjected to lower stresses, before forming.

10.8.2.1 Advantages of Tailored Blank Welding The advantages of tailored blanks over conventionally produced body-in-white components include the following:

1. Weight reduction, achieved by welding a thinner gage to a thicker gage material that provides reinforcement only where needed and also by elimination of reinforcement panels. The elimination of weld flanges that are normally required for resistance spot welding of separately formed parts also reduces weight.
2. Improved part functional performance. This results from improved structural rigidity and body accuracy or dimensional control due to a reduction in assembly variation as a result of a decrease in the number of stampings being assembled. This eliminates tolerance stackup, since tailored blanks join sheets before they are stamped, rather than welding them after being stamped.
3. Cost reduction through reduced stamping and assembly costs, and offal recovery. Offal is the scrap metal blank that falls off during sheet metal blanking from coil. Small pieces from the offal can be joined to form a functional single blank.
4. Reduced sealing needs.
5. Improved crash worthiness, since laser welds are stiffer than corresponding spot welds.

The advantages of tailored blank welding are further illustrated in Table 10.5, which compares two traditional methods of producing the body side frame with tailored blank welding. In Table 10.5, the separation (divided) method uses individual components (Figure 10.29), which are first formed, and then joined, usually by spot welding. In the integration (one-sheet) method, cuts are made in a large sheet panel to leave one large panel of the desired shape, which is then formed. In a sense, tailored blank welding combines the two approaches by first welding appropriate individual components to form the large panel of the desired shape (Figure 10.30).

10.8.2.2 Disadvantages of Tailored Blank Welding The disadvantages of tailored blank welding include the following:

1. Formability performance. Due to the difference in thickness and/or strength of the two sheets, forming of tailored blanks without producing defects is more difficult than traditional forming of single sheets.

Table 10.5 Comparison of two traditional methods of producing automotive body components with components produced by tailored blank welding (Table 1, Nakagawa et al., SAE Publication #930522, 1993).

	Divided method	One-sheet method	Tailored blanks
Appearance	Poor	Good	Good
Accuracy	Low	High	High
Yield	High (65%)	Low (40%)	High
Material flexibility	Selectable	Fixed	Selectable
Number of dies required	High	Low	Low

Source: Adapted from Nakagawa et al. (1993).

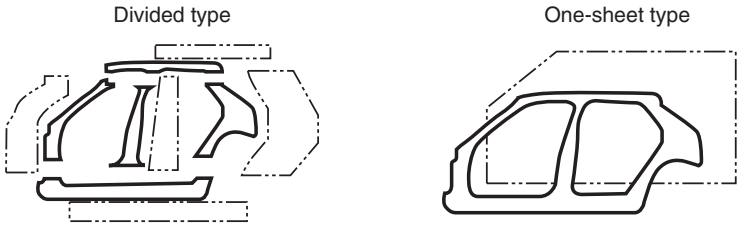


Figure 10.29 Illustration of the separation (divided) and integration (one-sheet) methods. Source: From Natsumi et al. (1991)/Inderscience Enterprises Ltd.

2. Required special edge preparation of the blanks. The gap between the blanks should not be more than 10% of the thickness of the thinner blank. This is often achieved by precision shearing, especially for long continuous welds of length greater than 1 m. Precision shearing may not be necessary for shorter welds, since they can often be processed with a die cut edge. The need for precision shearing, even for long welds, can also be eliminated by using filler wire or beam weaving. However, these techniques tend to slow the process and add to cost.
3. Potential for hot cracking in steels containing significant amounts of sulfur and phosphorus.
4. Potential for enhancing die wear. Since the weld bead is usually harder than the base metal, the welded sheet has a greater tendency to wear off die material than a single sheet that consists of the base metal.
5. Removal of protective coating in the heat-affected zone of the weld. Since the heat-affected zone size in laser welding is relatively small, the loss of coating may not be a significant factor.

10.8.2.3 Applications of Laser-Welded Tailored Blanks Even though there are a number of applications for tailored blanks in automotive body components, not all components can be candidates for tailored blanks. Some of the common production applications in the automotive industry are shown in Figure 10.30 and include center pillars, body side frames, upper and lower frame rails, roof reinforcement, rear deck lid, rear and front inner door panels, and floor panel. A typical vehicle may have about 20 components that are candidates for tailored blank welding. Multi-gage inner panels are considered to be among the tailor-welded blank applications that result in the highest benefits.

Some of the issues of concern that are often considered in setting up tailored blank applications include the welding costs, weld seam characteristics such as quality, reliability of the welding system, effect of welding on coatings, and weldment formability. Some of the important cost issues include cost of the welding process, work material, blanking, pallets, transportation, blank protection, inventory, and quality control. Critical factors that need consideration in application include the beam

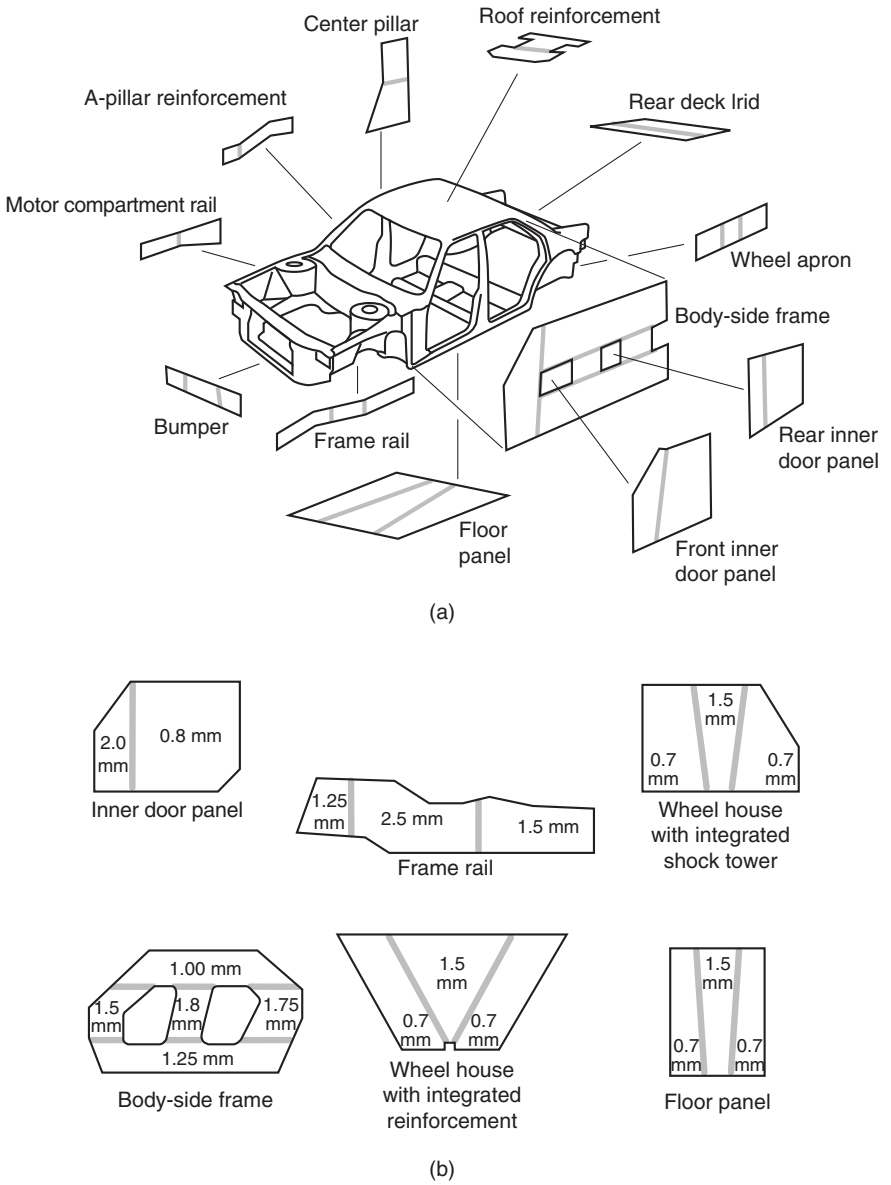


Figure 10.30 (a) Automotive body-in-white showing typical laser-welded tailored blanks. (b) Candidate laser-welded tailored blank components in an automotive body-in-white.

power and characteristics, gap width, edge preparation, and clamping. Processing speeds typically fall in the range of 4–10 m/min, depending on the laser power and sheet thickness.

The production system for laser-welded tailored blanks can be a fully automatic line with a number of workstations, or a manual single station operation. The latter would normally have a single welding system (which would include facilities for blank shearing, laser beam, and gantry), and two or more operators that perform the tasks of loading and unloading. The workstation may be either a single- or two-axis system. Single-axis systems are more suitable for in-line or long continuous welds which would normally be over a meter in length. Two-axis systems are appropriate for multiple,

short welds in any planar orientation. In either case, the overall blank size that can be processed is determined by the gantry and fixturing used.

Components of a full production system may include pallets, destacker, precision shearing for edge preparation, material handling, laser generator, inspection station, dimpling station for multi-gage blanks, and flipper when both left and right parts are made, for proper stacking of the parts on the pallet. Optional items may include blank washing, weld cleaning, oiling, and trimming. Some important factors to be considered in designing a production system include the following:

1. Sheet gage – single or multi-gage. This determines the type of clamping device to be used. Also, dimplers are required for the thinner sheet of multi-gage blanks, to ensure even stacking on the pallet.
2. Desired length of weld. This determines the gantry specifications.
3. Number of welds required per part.

10.8.2.4 Formability of Tailor-Welded Blanks Formability is a measure of the ease of forming a given material or body without failure. The formability of welded blanks is influenced by weld parameters such as the following:

1. Butt joint gap, which is the space between the two sheets (Figure 10.31a).
2. Beam misalignment, which is the offset of the centerline of the laser beam from the centerline of the gap (Figure 10.31b).
3. Shielding gas type and flow rate.
4. Focal point location.
5. Welding speed.
6. Laser power.

Some observations regarding the influence of process parameters on formability may be summarized as follows:

1. Formability increases with increasing welding speed, since the increased welding speed results in a reduced weld section, even though the hardness of the weld increases.

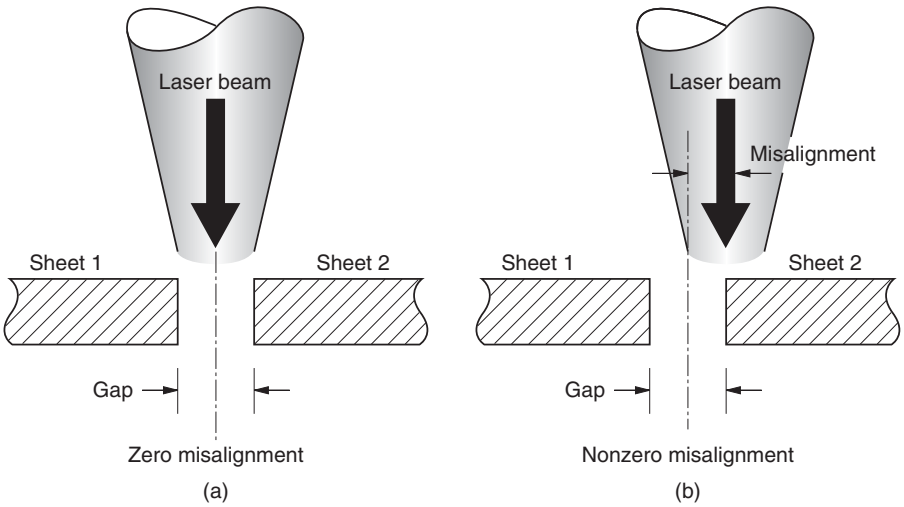


Figure 10.31 Illustration of (a) joint gap and (b) beam misalignment.

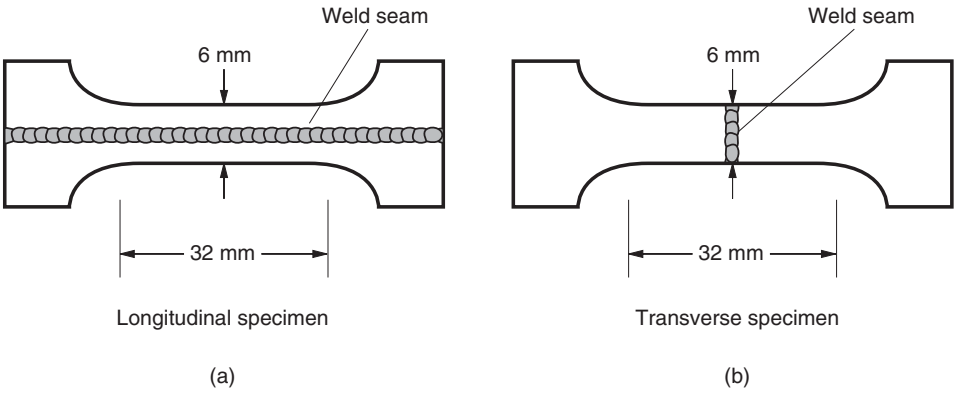


Figure 10.32 Tensile specimens of tailor welded blanks. (a) Longitudinal specimen. (b) Transverse specimen.

2. An increase in misalignment or joint gap reduces formability.
3. Misalignment or joint gap tend to promote high weld concavity, and these have relatively high failure rates. Generally, misalignment and joint gap have to be less than 0.05 mm to enhance quality and formability.

Uniaxial tension tests provide the simplest approach to assess the formability of tailor-welded blanks. It is more appropriate to use longitudinal specimen, that is, those cut along the weld with the welding line in the center of the specimen (Figure 10.32a), since these provide information on the mechanical properties of the welded blank. Transverse specimen (Figure 10.32b) do not provide adequate formability information, since deformation then occurs primarily in the base metal outside the weld. The higher yield strength of the weld then prevents plastic deformation in the weld. Transverse specimen thus primarily provide information on the weldment quality. It should be noted, however, that even for the longitudinal specimen, the tensile properties that are measured would depend on the ratio of the weld width to the entire tension specimen width. Thus, the specimen widths that are used, typically about 6 mm, are preferably smaller than the standard widths used for tensile tests.

When the direction of principal strain is in the direction of the weld, failure normally occurs in the weld, with the crack running transversely (Figure 10.33a). This form of failure may occur in welds made from materials with similar thicknesses/strengths. The other form of failure involves a crack that runs parallel to and is adjacent to the weld and occurs primarily in welds made from dissimilar thicknesses/strengths (Figure 10.33b). This normally occurs when the direction of principal strain is normal to the weld line, and fracture tends to occur in the lower strength or thinner material. Thus, optimum formability performance may be achieved by placing the weld line away from the major strain direction and/or away from the areas with high component strain level in the direction along the weld line, preferably perpendicular to the major strain direction.

10.8.2.5 Limiting Thickness or Strength Ratio When the thickness ratio or strength ratio of a tailor-welded blank exceeds a critical value, the material with the higher strength or thickness may be subjected to very little or no plastic deformation in the direction transverse to the weld direction, while the material of lower strength or thickness exceeds its forming limit near the weld. The limiting thickness ratio (LTR) (for same materials with different thicknesses) or limiting strength ratio (LSR) (for different materials with a prespecified thickness ratio) is the thickness or strength ratio at which one material just reaches initial yield strength when the other material reaches its forming limit. When this value is exceeded, failure may occur parallel to the weld, with cracking occurring on the

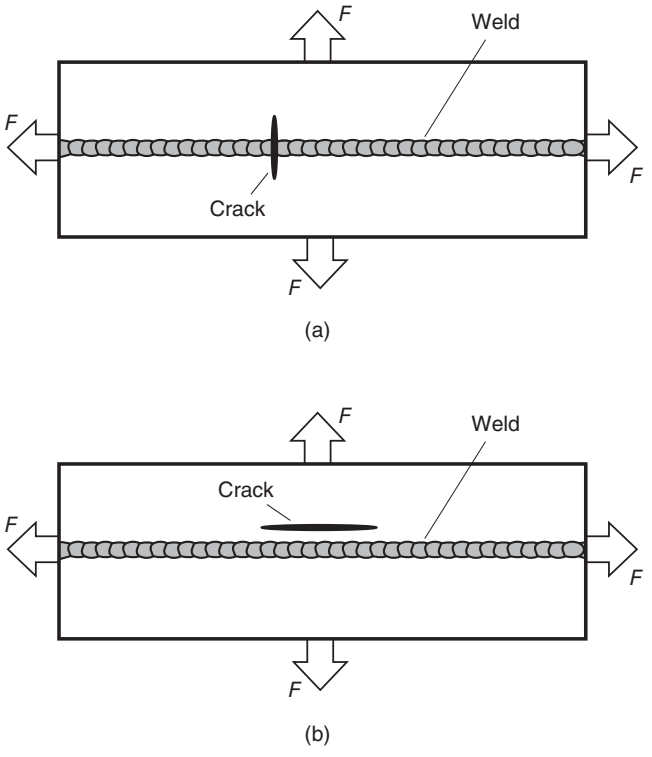


Figure 10.33 Cracking during forming of tailor welded blanks. (a) Crack running transverse to weld. (b) Crack running parallel to weld.

side of the material of lower thickness or strength. When the thickness and strength ratios are less than the LTR and LSR, respectively, plastic deformation would occur in both materials.

To develop an expression for the LTR or LSR, let us consider Figure 10.34, which shows two materials (1 and 2) that are joined by a laser weld. Neglecting friction effects, equilibrium in the transverse direction gives

$$F_1 = F_2 \tag{10.35a}$$

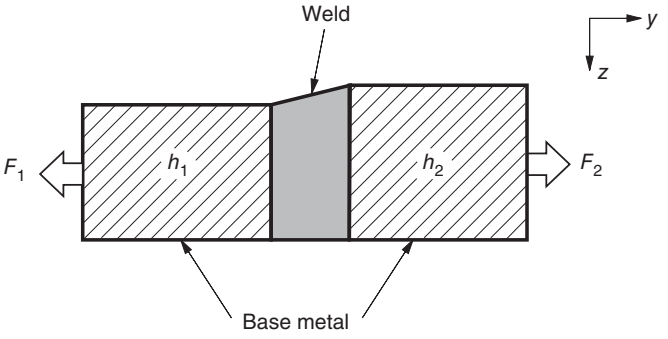


Figure 10.34 Force equilibrium in the transverse direction of a tailored weld.

where F is the force per unit width and the subscripts 1 and 2 represent the materials 1 and 2, respectively.

Eq. (10.35a) can be written as:

$$\sigma_1 h_1 = \sigma_2 h_2 \tag{10.35b}$$

where σ is the true stress and h is the instantaneous material thickness.

The true strain, ε_z, in the thickness direction is given by:

$$\varepsilon_z = \ln\left(\frac{h}{h_0}\right) \tag{10.36}$$

where h_0 is the original thickness.

Thus,

$$h = h_0 e^{\varepsilon_z} \tag{10.37}$$

Considering volume constancy during plastic deformation, we have

$$\varepsilon_x + \varepsilon_y + \varepsilon_z = 0 \tag{10.38}$$

where ε_x is the true strain in the welding direction (parallel to the weld in the sheet plane) and ε_y is the true strain in the transverse direction (normal to the weld in the sheet plane).

From Eqs. (10.35a,b), (10.37), and (10.38), we get

$$\sigma_1 h_{01} e^{-(\varepsilon_x)_1 - (\varepsilon_y)_1} = \sigma_2 h_{02} e^{-(\varepsilon_x)_2 - (\varepsilon_y)_2} \tag{10.39}$$

The strain component parallel to the weld (longitudinal strain) must be the same in both materials (1 and 2), that is,

$$(\varepsilon_x)_1 = (\varepsilon_x)_2 \tag{10.40}$$

Thus,

$$\sigma_1 h_{01} e^{-(\varepsilon_y)_1} = \sigma_2 h_{02} e^{-(\varepsilon_y)_2} \tag{10.41}$$

or

$$S_1 h_{01} = S_2 h_{02} \tag{10.42}$$

where S is the engineering stress.

If two sheets of the same material but different thicknesses are being welded, then from Eq. (10.42), the LTR can be obtained from

$$S_T h_{01} = S_y h_{02} \tag{10.43}$$

where S_T is the tensile strength of the material and S_y is the yield strength of the material.

The LTR is then given by:

$$\mathrm{LTR} = \text{limit of } \left(\frac{h_{02}}{h_{01}}\right) = \frac{S_T}{S_y} \tag{10.44}$$

If, on the other hand, the two sheets are made from different materials but have a prespecified thickness ratio, then Eq. (10.42) becomes

$$S_{T1} h_{01} = S_{y2} h_{02} \tag{10.45}$$

and the LSR is then given by:

$$\text{LSR} = \text{limit of } \left(\frac{S_{T1}}{S_{y2}}\right) = \frac{h_{02}}{h_{01}} \tag{10.46}$$

The thickness ratio $\frac{h_{02}}{h_{01}}$ normally has to be 1.25 or less, since the formability of the part reduces for thickness ratios above 1.25.

10.8.3 Laser Transmission Welding of Plastics

Laser transmission welding has evolved as an effective approach to joining plastics, especially thermoplastics. One of the workpieces to be joined is generally transparent to the laser beam, while the other workpiece is designed to have material properties that enable it to effectively absorb the beam. The two pieces are kept in intimate contact, usually by clamping. For the simple case of a lap joint, as illustrated in Figure 10.35, the laser beam is incident on and transmitted through the transparent plastic (top layer) to the interface between the two workpieces, where significant absorption of the beam occurs. This is made possible by either coating the interface with a highly absorbing material or making the lower workpiece absorbing by pigmentation, usually with carbon black. Additional absorption then occurs as the beam propagates further into the absorbing workpiece. The energy absorbed at the interface and in the absorbing workpiece heats up the interface and surrounding areas. Material in this region thus expands, increasing the interface pressure, and thus thermal contact conductance, which further enhances heat transfer to the transparent workpiece. The end result is melting of material in the region, accompanied by diffusion of the polymer molecules, thereby forming the joint on solidification.

Most thermoplastics are transparent to radiation in the near-infrared range (0.78–3 μm), more specifically up to 1.5 μm. Thus, Nd:YAG, diode, and some fiber lasers are very effective for laser transmission welding, enabling clear polymers to be welded to other polymers or to metals. A high-speed galvanometer scanner (wobble welding) may be used to control heat input and melting of the polymer at the interface where the beam is mostly absorbed.

The transparency depends on the extent of crystallinity of the thermoplastic, as the crystalline structure scatters part of the laser beam. The beam transmission (T_r) is also affected by fillers, colorants, reinforcement, and other modifiers, all of which result in additional diffuse scattering (multiple specular reflectance). This increases the effective optical path length of the transmitted beam and also deflects part of the beam from the incident beam direction. Analysis of transmission

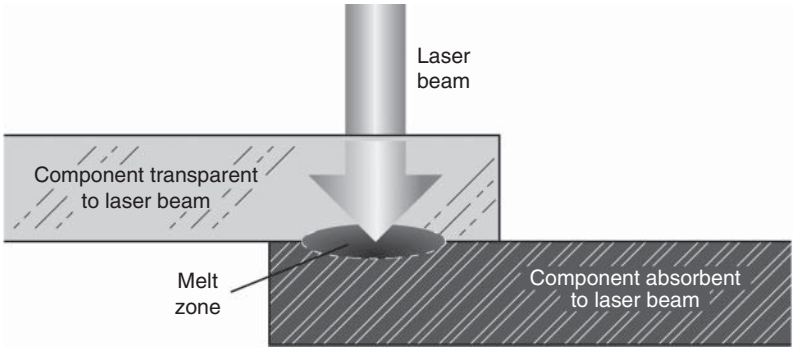


Figure 10.35 Schematic of the laser transmission welding process.

through the transparent material can therefore be quite complex but is often approximated using the Beer–Lambert law as:

$$T_r = 10^{-A_b} \tag{10.47}$$

where A_b is the absorbance. For low absorbance, the overall beam transmittance, T_{or}, can be approximated as:

$$T_{or} = 1 - A_b - R \tag{10.48}$$

where R is the reflectance.

The process can be implemented in different ways as follows:

1. The more traditional seam (contour) welding, where the heat source moves along the joint.
2. Simultaneous (spot) welding, where the laser beam covers the entire region to be joined, heating and bonding the two pieces together simultaneously.
3. Mask welding, where a mask is used to direct the laser beam only to regions that need to be joined.

The temperature distribution that results in the workpieces depends on the manner in which the laser beam is applied (contour, simultaneous, or mask), as well as the absorption method (interface coating or pigmentation). Increasing the amount of carbon black in the absorbing workpiece reduces the depth of molten region in that workpiece and an increase in the depth in the transparent workpiece, since most of the heat is then generated close to the interface.

The temperature distribution also depends on whether the workpieces are reinforced or not. The presence of fibers and their orientation affect the workpiece absorption and thermal conduction properties, and consequently, the weld bead geometry.

Typical processing parameters might consist of a 150 W CW power output that is focused using a 100–300 mm focal length lens, with speeds ranging from 450 to 4000 mm/min and pressures between 0.7 and 4.5 MPa. Excessive energy input may degrade the polymer.

Advantages of laser transmission welding include the following:

1. Ability to control energy deposition with high precision.
2. Noncontact process.
3. High production rates.
4. Suitability to sensitive components (for example, medical and electronic parts), since there is no vibration during the process.
5. Minimal heat-affected zone.

Application areas for laser transmission welding also include

1. Automotive industry
   - Fuel line components
   - Tail light assemblies.
2. Electrical components
   - Cell phone components – housing, connectors, and display windows.
3. Medical industry
   - Housings, valves, and optical lenses.
4. Aerospace industry
   - Aircraft interior components.

10.8.4 Laser Brazing

Laser brazing is simply another form of brazing where a laser is used as the heat source, instead of the traditional sources such as a furnace or gas torch. It is used for joining dissimilar materials which are not easily welded using conventional fusion welding processes. The advantage with the laser is the localized heat input, which minimizes the formation of brittle intermetallics. The amount of undesirable phases formed depends on the process time/temperature history, which may vary with location, depending on the workpiece geometries, and the type of filler material used, if any.

There are two basic forms of laser brazing. These are as follows:

1. Non-autogenous.
2. Autogenous.

10.8.4.1 Non-Autogenous Laser Brazing Non-autogenous laser brazing is essentially a laser-based variation of conventional brazing in that a filler metal is made to flow between faying surfaces.

The process has been used to braze a 5052 aluminum alloy to a 304 stainless steel, and also interstitial free steel, both of thickness 1.2 mm, using a diode laser with an aluminum filler metal, 1.2 mm in diameter, and Nocolock flux. Joint strengths obtained were as high as 80% the strength of the aluminum alloy, increasing with the laser power, up to an optimum value (at laser powers of 1200–1350 W), and then decreasing. Al–Fe intermetallic compounds are formed, but the amount depends on the laser power and other brazing conditions. So does the weld quality in general. The tendency to crack increases with the intermetallic layer thickness, which itself increases with laser power. The diode laser is effective because of its short wavelength and high absorptivity in metals. Since wettability of molten metal improves with increasing temperature, penetration is enhanced at higher powers. The braze wire feed speed is normally the same as the brazing speed. Laser beam size may range between 1 and 2 mm.

Non-autogenous laser brazing has also been used to join ceramic to steel. A major challenge to this process is the poor wettability of molten metals on ceramics. This is mitigated by using filler materials that contain Ti or Zr to improve wettability. Another challenge is the large differences in thermal expansion coefficient and elastic modulus of ceramics and metals. This can be addressed by placing a layer of ductile metal between the two workpieces, minimizing the residual stresses induced in the ceramic. For the specific case of aluminum oxide to steel, an AgCuTi-foil of thickness 50 μm was used as the braze filler metal. A CO_2 laser, defocused to a spot size of about 5 mm, was used as the heat source at power levels between 300 and 360 W and offset slightly toward the steel workpiece. The resulting joint strength though was lower than that of the original ceramic.

10.8.4.2 Autogenous Laser Brazing Autogenous laser brazing is effectively used to join dissimilar metals without using filler material. It involves the application of heat to the higher melting material, which transfers the thermal energy to the lower melting component by conduction. The interface is maintained at a temperature between the melting points of the two metals to be brazed.

The process has been effectively used to join a 6056 aluminum alloy to low-carbon steel using an Nd:YAG laser. First, Nocolok flux mixed with alcohol is precoated on the faying surfaces to break down any oxide film. The laser beam then irradiates the steel surface, which conducts the heat to the aluminum workpiece, resulting in melting of a thin layer of aluminum material (of the order of microns) at the interface, which subsequently solidifies to form a joint. Interface temperatures of about 800 °C have been found to result in good joints. Since the steel has a much higher melting temperature, it does not melt. Higher interface temperatures increase the thickness of the molten layer and its wettability (smaller wetting angle), as well as diffusion into the solid metal. Cracks

may form when the process results in an intermetallic layer of thickness greater than 10 μm, but that depends on the nature of intermetallics that are formed.

Titanium and aluminum alloys, specifically Ti6Al4V and A6061 of 2 mm thickness, have also been joined by fiber laser brazing without filler wire using a 4 kW laser at a speed of 4 m/min. For a butt joint, when the laser beam is offset a distance of 0.8 to 1.0 mm into the Al workpiece, there is no melting of the titanium alloy. The joint strength then increases as the offset distance increases, since the thickness of intermetallic $TiAl_3$ formed at the interface is then relatively small. When the offsetting distance is less than 0.8 mm, partial melting of the titanium alloy occurs, producing cracks and porosity caused by intermetallic compounds.

10.9 SUMMARY

Laser welding can be done using CO_2, Nd:YAG, and fiber lasers. The key parameters that influence the process are the beam power and characteristics, traverse speed, shielding gas, relative location of the focal point, and joint configuration. As an example, a 12 mm-thick carbon steel plate can be welded with a CO_2 laser using an output power of 9 kW at a speed of 0.8 m/min, or a 6 mm-thick carbon steel plate can be welded with a CO_2 laser using an output power of 25 kW at a speed of 10 m/min. As the welding speed increases, the depth of penetration for a given power decreases.

A Gaussian beam mode results in a greater depth-to-width ratio than higher-order modes. However, the larger spot size of multimode beams makes them more suited for butt welding applications, since it reduces the tight fit-up requirements. The effect of polarization on the process depends on the material and welding conditions. For some steels, the depth of penetration is increased when the direction of polarization is aligned with the welding direction, and when the welding speed is above a critical value.

Both pulsed and CW laser beams can be used for welding, with pulsed beams being more suited for spot welding, while CW beams are more suited for continuous welds. Pulsed beams can also be used for continuous welds if the pulses are made to overlap. With high-power beams, a portion of the laser energy is expended in ionizing part of the shielding gas and/or vapor to form a plasma. This plasma may absorb part of the laser beam and may also induce an optical lensing effect that changes the effective focusing of the laser beam. Shielding gases are used to blow away the vapor and plasma formed, provide a protective environment for the weld pool, and protect the focusing lens. Common shielding gases used are argon and helium. Argon is more commonly used for low- to medium-power laser welding. It is also generally preferred to have the focal point either on the workpiece surface or just below it.

Laser welding may occur in the conduction mode (normally used for foils and thin sheets), or keyhole mode (for relatively thick sections). In conduction mode welding, the laser power is first deposited on the surface and then transferred by conduction to the surroundings. Absorption in the keyhole occurs by inverse bremsstrahlung of the plasma and by Fresnel absorption (direct deposition) at the cavity walls.

Low-carbon steels and some stainless steels are easily welded with a laser, while brass, silver, and galvanized steel are more difficult to weld. Common defects that affect product quality include porosity, cracking, and humping. A dual-beam system can be used either to control the cooling rates and therefore the microstructure in the solid material (by preheating and/or post-weld heat treating), or to control flow in the weld pool. Combining a laser beam with an arc welder increases the depth of penetration obtained compared to that for either process alone.

With wobble welding, where the laser beam is scanned with a specific geometry along the joint, the weld penetration and width can be controlled, and the formation of discontinuities minimized.

RLW also enables the welding process to be accomplished using long focal length optics and is useful in applications where joint accessibility is difficult.

Lasers are effectively used in microwelding and also in tailored blank welding, which involves welding of sheet metal blanks of either different thicknesses (gages), coatings (bare or galvanized), or materials (grades). In recent years, lasers have seen increasing use in welding of plastics where one of the workpieces to be welded is made transparent to the laser beam, while the other is absorbing. The energy accumulated by the absorbing layer heats up the workpieces at the interface. Nd:YAG, diode, and fiber lasers are most suited to this type of welding, referred to as laser transmission welding. Lasers have also been increasingly used for brazing, especially of dissimilar metals, because of the localized heat input, which minimizes the formation of brittle intermetallics. The process can be either autogenous or non-autogenous. Non-autogenous welding is preferred when there is poor wettability between the two materials. This is mitigated by using filler materials that improve wettability.

APPENDIX 10.A

List of symbols used in the chapter

Symbol	Parameter	Units
A_b	Absorbance	—
d	Inter-beam spacing	mm
D	Maximum width of molten zone	μm
F	Force per unit width	N/m
$\mathbf{F_b} = -\rho_l \mathbf{g} \beta_T (T - T_0)$	Body force vector	N/m^3
$\mathbf{F_i}$	Body force components (F_x, F_y, F_z)	N/m^3
h_0	Original thickness	mm
$\mathbf{I_0}$	Intensity of incident laser beam	W/m^2
l	Molten pool length	μm
l_f	Molten pool length ahead of keyhole	μm
LTR	Limiting thickness ratio	—
LSR	Limiting strength ratio	—
P_{ab}	Ablation pressure	Pa
P_{hd}	Hydrodynamic pressure	Pa
P_{hs}	Hydrostatic pressure	Pa
P_v	Excess pressure of vapor in keyhole	Pa
P_γ	Pressure due to surface tension	Pa
q_1	Heat flux (power) input to the workpiece of major heat source	W
q_2	Heat flux (power) input to the workpiece of minor heat source	W
r_k	Keyhole radius	μm
S	Engineering stress	Pa (N/m^2)
S_T	Tensile strength	Pa (N/m^2)
S_y	Yield strength	Pa (N/m^2)
t_D	Time it takes for melt to attain width D	s
$\mathbf{t}$	Unit tangential vector	—

Symbol	Parameter	Units
T_1	Absolute temperature of major heat source	K
T_2	Absolute temperature of minor heat source	K
T_i	Absolute temperature of ionized electrons	K
T_{or}	Overall beam transmittance	—
T_r	transmittance	—
$\mathbf{U_{cs}}$	Average speed at which heat is conducted sideways	m/s
$\mathbf{u_{il}}$	Velocity vector at solid–liquid interface	m/s
$\mathbf{u_l}$	Velocity vector for the liquid phase	m/s
$\mathbf{u_v}$	Velocity vector at liquid–vapor interface	m/s
$\mathbf{u_{iv}}$	Velocity vector at free surface (vaporization interface)	m/s
u_{vt}	Average velocity of vapor flow transversely across the cavity	m/s
$\xi' = \xi - d$	Coordinate	—
w_1	Average sideways width of molten pool	μm
z_i	Ionized electrons per ion	—
α_p	Absorption coefficient of plasma	/m
β_v	Fraction of material transported across cavity as vapor	—
∇_{lv}	Surface gradient at the liquid–vapor interface	—
η_g	Generator efficiency	%
η_p	Process efficiency	%
η_{th}	Transmission/heat transfer efficiency	%
γ_{lv}	Surface tension at liquid–vapor interface	N/m
ρ_e	Electron density	/mm^3
ρ_i	Density of ions	/mm^3
ρ_v	Density of the vapor at temperature T_v	kg/m^3
τ	Stress tensor at liquid–vapor interface	Pa
θ	Angle of incidence of the laser beam on the keyhole wall at any location, measured from the Surface normal	°
Θ	Temperature	K

PROBLEMS

10.1. Determine the welding speed above which heating in front of the molten pool is adiabatic, if the material being welded is a mild steel, and the keyhole diameter is 1 mm

$$u_x > \frac{\kappa_s}{r_k}$$

10.2. An Nd:YAG laser with two output beams, each of power 3.5 kW is used to weld a pearlitic steel, with one beam being used for preheating to reduce the cooling rates. Determine the maximum welding speed, u_x, that can be used to avoid martensite formation, if the spacing between the two beams is $d = 5$ mm, and 90% of each beam is absorbed by the workpiece. Further assume that the density of molten steel is $\rho_l = 7407\,\text{kg/m}^3$, and that the critical cooling rate to avoid martensitic transformation in pearlitic steels is about 150 °C/s at a temperature of 550 °C.

10.3. For the conditions in Problem 10.2, what would be the maximum welding speed if a single beam of power 3.5 kW was used? Assume an ambient temperature of 0 °C.

10.4. If the beam quality of the Nd:YAG laser used in Problem 10.3 is $M^2 = 20$, and the distance between the cavity mirrors is 400 mm, determine the radius, w_f, of the focused beam. Assume the focusing lens is positioned at a distance 2.5 m from the beam waist and that the lens focal length is 200 mm.

10.5. Using the welding speed, u_x, obtained from Problem 10.3, and assuming keyhole mode welding with the keyhole radius, r_k, being equal to the radius, w_f, of the focused beam, and further that the ratio of solid density to vapor density is approximately 10^4, estimate

(1) The weld pool length ahead of the keyhole, l_f.
(2) The average side width of the molten pool, w_l.
(3) The length of the molten pool behind the cavity, l.

Comment on the results obtained.

10.6. A thin aluminum plate is welded using a laser beam with the following conditions
Power input = 8 kW
Plate thickness = 3 mm
Welding speed = 75 mm/s
Initial plate temperature = 25 °C
Heat transfer efficiency = 0.6.
Calculate the temperature on the top surface of the plate, 4.0 mm behind the laser beam and 1.5 mm to one side of it.

10.7. For the material and conditions in Problem 10.6, determine the cooling rate experienced at the projection of that point directly behind the heat source.

10.8. Given that

$$\beta_v = \frac{1}{\left[1 + \frac{l_f}{r_k}\left(\frac{2\rho_l}{\rho_v}\right)^{1/2}\right]}$$

Show that

(a) $\beta_v \propto u_x$,
(b) $u_{vt} \propto u_x^2$, and
(c) $P \propto u_x^4$.

10.9. Show that Eq. (10.10) given by:

$$(c_{vol}T_m + L_{mv})u_x = k_l\frac{T_v - T_m}{l_f}\left(1 - \frac{u_x l_f}{\kappa_l}\right)$$

reduces to the form:

$$l_f = \frac{\kappa_l}{u_x}\left(\frac{T_v - T_m}{T_v + L_{mv}/c_{vol}}\right)$$

Hint: Note that $k_l = c_{vol}\kappa_l$.

10.10. Show that the following equation

$$w_l^2 = \frac{\kappa_l}{u_x}r_k\left(\frac{T_v - T_m}{(T_m + T_v)/2 + L_{mv}/c_{vol}}\right)$$

reduces to the form:

$$w_l \approx (l_f r_k)^{1/2}$$

Hint: Compare with Eq. (10.11), and ignore the small difference in the denominator.

10.11. Why is it more feasible to weld dissimilar materials using laser welding than with conventional arc welding?

10.12. Discuss the relevance of the derivation leading up to Eq. (10.42).

10.13. Show that the maximum value of the absorption coefficient occurs at high angles of incidence, almost glancing incidence.

10.14. A 5 kW CO_2 laser is used to weld a 3 mm-thick steel plate at a speed of 10 m/min. If the radius of the focused beam, w_f, is 75 μm and is half the keyhole radius, r_k, calculate:

(a) The weld pool length in front of the keyhole, l_f.

(b) The fraction of material transported across the cavity as vapor, β_v.

(c) The length, l, of the molten pool (from the keyhole center to the tip of the molten pool)
Adiabatic conditions may be assumed to exist ahead of the weld pool, and further that the ratio of molten pool density to vapor density is 10^4. The properties of the steel plate are as follows:
Density of molten metal, $\rho_l = 7407\ \text{kg/m}^3$
Density of solid metal, $\rho_s = 7870\ \text{kg/m}^3$
Specific heat, $c_p = 452\ \text{J/kgK}$
Thermal conductivity, $k = 73\ \text{W/mK}$.
Vaporization temperature, $T_v = 2870\ ^\circ\text{C}$
Melting temperature, $T_m = 1538\ ^\circ\text{C}$
Latent heat of melting, $L_m = 247\ \text{kJ/kg}$
Latent heat of vaporization, $L_v = 7018\ \text{kJ/kg}$

10.15. Starting with Bernoulli's equation, show that the pressure change that results from flow of vaporized material from the front of the cavity is given (neglecting viscosity effects) by

$$P = {}^1\!/_2(\beta_v \rho_l u_x) u_{vt}$$

10.16. An Nd:YAG laser of power 3.5 kW is used to weld a pearlitic steel at a speed $u_x = 120$ mm/s. Assuming keyhole mode welding with the keyhole radius, r_k, being equal to three times the radius, $w_f = 50$ μm, of the focused beam, and further that the ratio of solid density to vapor density is approximately 10^4, estimate:

(a) The weld pool length ahead of the keyhole, l_f.

(b) The average side width of the molten pool, w_l.

(c) The length of the molten pool behind the cavity, l.

Comment on the results obtained.
You may further assume that the density of molten steel is $\rho_l = 7407\ \text{kg/m}^3$, and that the properies of steel are the same as those for iron.

10.17. A CO_2 laser is split into two beams of equal power and used for butt welding two thick plates of aluminum, each 5 mm in thickness. If the maximum cooling rate is about 1000 °C/s at a temperature of 450 °C which occurs at a distance of 15 mm behind the major source, what is the welding speed for the process? Assume the distance between the major and minor heat sources is $d = 4$ mm.

BIBLIOGRAPHY

Abilsiitov, G.A. and Velikhov, E.P. (1984). Application of CO_2 lasers in mechanical engineering technology in the USSR. *Optics and Laser Technology* 16: 30–36.

Bappa Acherjee, B., Kuar, A.S., Mitra, S., and Misra, D. (2012). Effect of carbon black on temperature field and weld profile during laser transmission welding of polymers: A FEM study. *Optics & Laser Technology* 44: 514–521.

Akhter, R., Davis, M., Dowden, J. et al. (1989). A method for calculating the fused zone profile of laser keyhole welds. *Journal of Physics D: Applied Physics* 21: 23–28.

Albright, C.E. and Chiang, S. (1988). High speed laser welding discontinuities, *Proc., 7th Int. Conf. on Applications of Lasers and Electro Optics, ICALEO'88*, Santa Clara Calif. pp. 207–213.

Arata, Y. and Nabegata, E. (1978). Tandem electron beam welding (Report-I). *Transactions of Japan Welding Research Institute* 7 (1): 101–109.

Arata, Y., Nabegata, E., and Iwamoto, N. (1978). Tandem electron beam welding (Report-II). *Transactions of Japan Welding Research Institute* 7 (2): 85–95.

Banas, C.M. (1992). Twin-spot lasers weld stainless tubing. *MAN* 14–16.

Bayshore, J.K., Williamson, M.S., Adonyi, Y., and Milian, J.L. (1995). Laser beam welding and formability of tailored blanks. *Welding Journal* 74: 345s–352s.

Becker, F. and Potente, H. (2002). A step towards understanding the heating phase of laser transmission welding in polymers. *Polymer Engineering and Science* 42 (2): 365–374.

Beyer, E., Behler, K., and Herziger, G. (1988). Plasma absorption effects in welding with CO_2 lasers. *SPIE, High Power Lasers and Laser Machining Technology* 1020: 84–95.

Chan, C., Mazumder, J., and Chen, M.M. (1984). A two-dimensional transient model for convection in laser melted pool. *Metallurgical Transactions A* 15A: 2175–2184.

Chan, C. and Mazumder, J. (1987). One-dimensional steady-state model for damage by vaporization and liquid expulsion due to laser-material interaction. *Journal of Applied Physics* 62: 4579–4586.

Chande, T. and Mazumder, J. (1984). Estimating effects of processing conditions and variable properties upon pool shape, cooling rates, and absorption coefficient in laser welding. *Journal of Applied Physics* 56: 1981–1986.

Chen, M., Zak, G., and Bates, P.J. (2011). Effect of carbon black on light transmission in laser welding of thermoplastics. *Journal of Materials Processing Technology* 211: 43–47.

Chen, T.-C. and Kannatey-Asibu, E. Jr. (1995). Dual beam laser systems and their impact on conduction mode weld pool convection and surface deformation. *Transactions of NAMRI/SME* 23: 151–156.

Chen, T.-C. and Kannatey-Asibu, E. Jr. (1996). Convection pattern and weld pool shape during conduction-mode dual beam laser welding. *Transactions of NAMRI/SME* 24: 259–265.

Cline, H.E. and Anthony, T.R. (1977). Heat treating and melting material with a scanning laser or electron beam. *Journal of Applied Physics* 48 (9): 3895–3900.

Conti, R.J. (1969). Carbon dioxide laser welding. *Welding Journal* 48: 800–806.

Dawes, C. (1992). *Laser Welding: A Practical Guide*. Cambridge, England: Abington Publishing.

DebRoy, T., Basu, S., and Mundra, K. (1991). Probing laser induced metal vaporization by gas dynamics and liquid pool transport phenomena. *Journal of Applied Physics* 70: 1313–1319.

Dharmendra, C., Rao, K.P., Wilden, J., and Reich, S. (2011). Study on laser welding–brazing of zinc coated steel to aluminum alloy with a zinc based filler. *Materials Science and Engineering A* 528: 1497–1503.

Ding, J., Li, F., Qu, F. et al. (2005). Investigation on laser brazing AA6056 Al alloy to XC 18 low-carbon steel. *Chinese Optics Letters* 3 (1): 31–34.

Douglas, D.M., Mazumder, J., and Nagarathnam, K. (1995). Laser welding of Al 6061-T6, *4th International Conference on Trends in Welding Research*, Gatlinburg, TN, June 5–8, 1995.

Douglas, D.M. and Mazumder, J. (1996). Mechanical properties of laser welded aluminum alloys. *ICALEO* 31–38.

Dowden, J., Postacioglu, N., Davis, M., and Kapadia, P. (1987). A keyhole model in penetration welding with a laser. *Journal of Physics D: Applied Physics* 20: 36–44.

Dowden, J., Kapadia, P., and Postacioglu, N. (1989). An analysis of the laser-plasma interaction in laser keyhole welding. *Journal of Physics D: Applied Physics* 22: 741–749.

Dowden, J., Chang, W.S., Kapadia, P., and Strange, C. (1991). Dynamics of the vapour flow in the keyhole in penetration welding with a laser at medium welding speeds. *Journal of Physics D: Applied Physics* 24: 519–532.

Dudas, J.H. and Collins, F.R. (1966). Preventing weld cracks in high-strength aluminum alloys. *Welding Journal* 45: 241s–249s.

Duley, W.W. (1982). *Laser Processing and Analysis of Materials*. New York: Plenum Press.

Eisenmenger, M., Bhatt, K.K., and Shi, M.F. (1995). Influence of laser welding parameters on formability and robustness of blank manufacturing: An application to a body side frame, SAE Technical Paper #950922, pp. 171–182.

Elmer, J.W., Giedt, W.H., and Eagar, T.W. (1990). The transition from shallow to deep penetration during electron beam welding. *Welding Journal* 69: 167s–176s.

Fairbanks, R.H. and Adams, C.M. (1964). Laser beam fusion welding. *Welding Journal* 43: 97s–102s.

Funk, M., Kohler, U., Behler, K., and Beyer, E. (1989). Welding of steel with a CO_2 laser of 20 kW. *SPIE, High Power Lasers and Laser Machining Technology* 1132: 174–180.

Fysikopoulosa, A., Pastrasa, G., Stavridisa, J. et al. (2016). On the performance evaluation of remote laser welding process: An automotive case study. *Procedia CIRP* 41: 969–974.

Gatzweiler, W., Maischner, D., Faber, F.J. et al. (1989). The expansion of the plasma is examined by using different positions of the slit of the streak. *SPIE, High Power Lasers and Laser Machining Technology* 1132: 157.

Grewell, D., Rooney, P., and Kagan, V.A. (2004). Relationship between optical properties and optimized processing parameters for through-transmission laser welding of thermoplastics. *Journal of Reinforced Plastics and Composites* 23 (3): 239–246.

Grezev, A.N., Grigor'yants, A.G., Fedorov, V.G., and Ivanov, V.V. (1984). The structure and mechanical properties of laser welded joints between dissimilar metals. *Automatic Welding* 63: 29–31.

Harry, J.E. (1974). *Industrial Lasers and their Applications*. London: McGraw Hill.

Irving, B. (1991). Blank welding forces automakers to sit up and take notice. *Welding Journal* 70: 39–45.

Irving, B. (1994). Automotive engineers plunge into tomorrow's joining problems. *Welding Journal* 73: 47–50.

Irving, B. and Baron, J. (1994). Mash-seam resistance welding fights it out with the laser beam. *Welding Journal* 73: 33–39.

Irving, B. (1995). Welding tailored blanks is hot issue for automakers. *Welding Journal* 74: 49–52.

Iwase, T., Shibata, K., Sakamoto, H. et al. (2000). Real time X-ray observation of dual focus beam welding of aluminum alloys. *ICALEO* C26–C34.

Jaeschke, P., Herzog, D., Haferkamp, H. et al. (2010). Laser transmission welding of high performance polymers and reinforced composites – A fundamental study. *Journal of Reinforced Plastics and Composites* https://doi.org/10.1177/0731684410365365.

Kagan, V.A., Bray, R.G., and Kuhn, W.P. (2002). Laser transmission welding of semi-crystalline thermoplastics — Part I: Optical characterization of nylon based plastics. *Journal of Reinforced Plastics and Composites* 21 (12): 1101–1122.

Kagan, V.A. and Pinho, G.P. (2004). Laser transmission welding of semi-crystalline thermoplastics — Part II: Analysis of mechanical performance of welded nylon. *Journal of Reinforced Plastics and Composites* 23 (1): 95–106.

Kangastupa, J. (2018). Fiber laser enables successful welding of difficult materials. *Industrial Laser Solutions* 24–25.

Kancharla, V., Mendes, M., Grupp, M., and Baird, B. (2018). Recent advances in fiber laser welding. *Industrial Laser Solutions* 11–16.

Kannatey-Asibu, E. Jr. (1991). Thermal aspects of the split-beam laser welding concept. *ASME Journal of Engineering Materials and Technology* 113: 215–221.

Kar, A. and Mazumder, J. (1995). Mathematical modeling of keyhole laser welding. *Journal of Applied Physics* 78 (11): 6353–6359.

Katayama, S. (2013). *Handbook of Laser Welding Technologies*. Cambridge, UK: Woodhead Publishing Limited.

Kim, C.-J., Kauh, S., Ro, S.T., and Lee, J.S. (1994). Parametric study of the two-dimensional keyhole model for high power density welding processes. *ASME Journal of Heat Transfer* 116: 209–214.

Klemens, P.G. (1976). Heat balance and flow conditions for electron beam and laser welding. *Journal of Applied Physics* 47 (5): 2165–2174.

Klingbeil, K. (2006). What you need to know about remote laser welding. *Welding Journal* 85 (8): 44–46.

Liu, Y.-N. and Kannatey-Asibu, E. Jr. (1993). Laser beam welding with simultaneous Gaussian laser preheating. *ASME Journal of Heat Transfer* 115: 34–41.

Liu, Y.-N. and Kannatey-Asibu, E. Jr. (1998). Finite element analysis of heat flow in dual-beam laser welded tailored blanks. *ASME Journal of Manufacturing Science and Engineering* 120: 272–278.

Locke, E.V., Hoag, E.D., and Hella, R.A. (1972). Deep penetration welding with high power CO_2 lasers. *IEEE Journal of Quantum Electronics* QE-8, 2: 132–135.

Lowry, J.F., Fink, J.H., and Schumacher, B.W. (1976). A major advance in high-power electron-beam welding in air. *Journal of Applied Physics* 47: 95–105.

Luxon, J.T. and Parker, D.E. (1985). *Industial Lasers and Their Applications*. Englewood Cliffs, New Jersey: Prentice Hall.

Magee, K.H., Merchant, V.E., and Hyatt, C.V. (1990). Laser assisted gas metal arc weld characteristics. *ICALEO* 382–399.

Markovit, T., Takács, J., Lovas, A., and Belt, J. (2003). Laser brazing of aluminium. *Journal of Materials Processing Technology* 143–144: 651–655.

Mathieu, A., Pontevicci, S., Viala, J.-C. et al. (2006). Laser brazing of a steel/aluminium assembly with hot filler wire (88% Al, 12% Si). *Materials Science and Engineering A* 435–436: 19–28.

Mathieu, A., Shabadi, R., Deschamps, A. et al. (2007). Dissimilar material joining using laser (aluminum to steel using zinc-based filler wire). *Optics & Laser Technology* 39: 652–661.

Matsuda, J., Utsumi, A., Katsumura, M. et al. (1988). TIG or MIG arc augmented laser welding of thick mild steel plate. *Joining & Materials* 1: 38, 31–41, 34.

Matsunawa, A. and Semak, V. (1997). The simulation of front keyhole wall dynamics during laser welding. *Journal of Physics D: Applied Physics* 30: 798–809.

Mayboudi, L.S., Birk, A.M., Zak, G., and Bates, P.J. (2007). Laser transmission welding of a lap-joint: Thermal imaging observations and three–dimensional finite element modeling. *ASME Journal of Heat Transfer* 129: 1177–1186.

Mazumder, J. and Steen, W.M. (1980). Heat transfer model for CW material processing. *Journal of Applied Physics* 51 (2): 941–947.

Mazumder, J. (1987). An overview of melt dynamics in laser processing. *High Power Lasers* 801: 228–241.

Metzbower, E.A. (1990). Laser beam welding: thermal profiles and HAZ hardness. *Welding Journal* 69: 272s–278s.

Mombo-Caristan, J.-C., Lobring, V., Prange, W., and Frings, A. (1993). Tailored welded blanks: A new alternative in automobile body design. In: *The Industrial Laser Handbook*, 1992–1993 edition, 89–102.

Nagayama, H., Mizutani, M., Kawahito, Y., and Katayama, S. (2007). Weld penetration and welding phenomena of aluminum alloy with high-power fiber laser. *ICALEO* 2007: 513.

Nakagawa, N., Ikura, S., Natsumi, F., and Iwata, N. (1993). Finite element simulation of stamping a laser-welded blank, SAE Technical Paper #930522, pp. 189–197.

Natsumi, F., Ikemoto, K., Sugiura, H. et al. (1991). Laser welding technology for joining different sheet metals for one-piece stamping. *JSAE Review* 12 (3): 58–63.

Prabhakara, R., Kontopoulou, M., and Zak, G. (2006). Contour laser – Laser-transmission welding of glass reinforced nylon 6. *Journal of Thermoplastic Composite Materials* 19: 427–439.

Rayes, M.E., Walz, C., and Sepold, G. (2004). The influence of various hybrid welding parameters on bead geometry. *Welding Journal* 83: 147s–153s.

Rofin-Sinar, *Technical Note* - CO_2 *Laser Welding*, Rofin Sinar, Plymouth, Michigan.

Rohde, M., Sudmeyer, I., Urbanek, A., and Torge, M. (2009). Joining of alumina and steel by a laser supported brazing process. *Ceramics International* 35: 333–337.

Saida, K., Song, W., and Nishimoto, K. (2005). Diode laser brazing of aluminium alloy to steels with aluminium filler metal. *Science and Technology of Welding and Joining* 10 (2): 227–235.

Shi, M.F., Pickett, K.M., and Bhatt, K.K. (1993). Formability issues in the application of tailor welded blank sheets. In: *SAE Technical Paper #930278*, 27–35.

Song, Z., Nakata, K., Wu, A., and Liao, J. (2013). Interfacial microstructure and mechanical property of Ti6Al4V/A6061 dissimilar joint by direct laser brazing without filler metal and groove. *Materials Science and Engineering A* 560: 111–120.

Steen, W.M. and Eboo, M. (1979). Arc augmented laser-welding. *Metal Construction* 11 (7): 332–335.

Steen, W.M. (1980). Arc augmented laser processing of materials. *Journal of Applied Physics* 51 (11): 5636–5641.

Steen, W.M. (2003). *Laser Material Processing*. Berlin: Springer-Verlag.

Swift-Hook, D.T. and Gick, A.E.F. (1973). Penetration welding with lasers. *Welding Journal* 52: 492s–499s.

Thompson, R. (1993). The LDH test to evaluate sheet metal formability - Final report of the LDH committee of the North American Deep Drawing Research Group. In: *Sheet Metal and Stamping Symposium, SAE Technical Paper #930815*, 291–301.

Utilase, I. (1994). *Eliminate Costs with Laser Welded Blanks*. Leaflet: Private Publication.

Van de Ven, J.D. and Erdman, A.G. (2007). Laser transmission welding of thermoplastics — Part I: Temperature and pressure modeling. *ASME Journal of Manufacturing Science and Engineering* 129: 849–858.

Voldrich, C.B. (1947). Cold cracking in the heat-affected zone. *Welding Journal* 26: 152s–169s.

Wilson, J. and Hawkes, J.F.B. (1987). *Lasers - Principles and Applications*. New York: Prentice Hall International Series in Optoelectronics.

Xue, Z., Hu, S., Shen, J. et al. (2014). Microstructure characterization and mechanical properties of laser-welded copper and aluminum lap joint. *Journal of Laser Applications* 26: 012002. https://doi.org/10.2351/1.4824778.

Yessik, M.J. (1978). Laser material processing. *Optical Engineering* 17 (3): 202–209.

Zacharia, T., David, S.A., Vitek, J.M., and Debroy, T. (1989). Weld pool development during GTA and laser beam welding of type 304 stainless steel, Parts 1 and 2. *Welding Journal* 68, 499s–509s, 510s–519s.

11 Laser Surface Modification

Laser surface modification is used to change the surface composition, microstructure, or geometry of a material to give it certain desired properties. This may involve hardening the surface to increase its resistance to wear, or inducing compressive residual stresses in the surface layers to enhance fatigue life.

There are six major forms of laser surface modification. These are the following:

1. Laser surface heat treatment.
2. Laser surface melting (skin melting or glazing).
3. Laser direct metal deposition (cladding, alloying, and hard coating).
4. Laser physical vapor deposition.
5. Laser shock peening.
6. Laser surface texturing.

In laser surface heat treatment, no melting takes place, while in laser surface melting, a thin surface layer of the workpiece is melted, and as a result of the rapid quenching, may form new structures which are harder. In laser deposition, a second material is applied to the surface and melted by the laser beam either to alloy with a thin surface layer of the base material or to bond to the surface. However, in laser physical vapor deposition, the second material is applied in vapor form, and no melting takes place. Laser shock peening, however, generates shock waves which induce compressive residual stresses on the surface of the material. Surface texturing is a means of modifying the geometric features of a surface to impart desired properties to it. These forms of surface modification are discussed in greater detail in the following sections according to the listed order.

11.1 LASER SURFACE HEAT TREATMENT

Surface heat treatment normally involves exposing the surface of a material to a thermal cycle of rapid heating and cooling such that the surface layers in the case of steels, for example, are first austenitized, and then quenched, to induce martensitic transformation. The process does not involve melting, and the transformations occur in the solid state.

The hardness of the martensite formed when steel is heat treated enhances the wear resistance of the surface. Examples of components that are hardened by laser heat treatment include laser-hardened wear tracks for power steering housing, cam shafts, automobile valve guides and seats, gear teeth, diesel cylinder liner bores, surface hardening of cylinder head units in aluminum car engines, hard-facing of car distribution shafts, and surface hardening of mill rollers.

Principles of Laser Materials Processing: Developments and Applications, Second Edition. Elijah Kannatey-Asibu, Jr.
© 2023 John Wiley & Sons, Inc. Published 2023 by John Wiley & Sons, Inc.
Companion Website: www.wiley.com/go/Kannatey-Asibu/PrinicplesLaserMaterialsProcessing

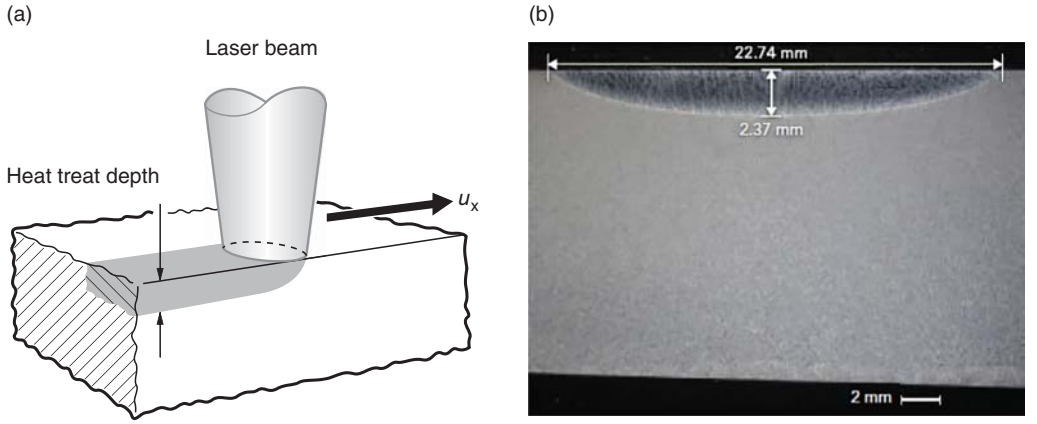


Figure 11.1 Laser surface heat treatment. (a) Set-up. Source: Adapted from Mazumder (1983). (b) Cross-section of S7140 alloyed steel after heat treatment. Source: Jonnalagadda et al. (2019)/Endeavor Business Media, LLC.

The setup for laser surface heat treatment is illustrated schematically in Figure 11.1. Rapid cooling results from mass or self-quenching into the bulk surrounding material and may result in martensitic transformation even in some low carbon steels. The cooling rates for laser heat treatment can be as high as 10^7 °C/s (compare with cooling rates of about 300 °C/s for arc welding) and produce relatively short austenitizing cycles (on the order of 0.01–1.0 second).

The essential steps involved in laser hardening normally comprise the following:

1. Cleaning: to remove surface contaminants such as oil.
2. Masking: to limit hardening to desired areas only. This step is not always necessary.
3. Coating: to increase beam absorption.
4. Hardening: the actual laser treatment.
5. Tempering: to reduce crack susceptibility of the fully martensitic case.
6. Cleaning: to remove any remaining coating.
7. Inspection: to ensure no defective parts are produced.

In this section, we first discuss criteria that are necessary for successful laser surface treatment. This is followed by a discussion on the key process parameters. A brief overview of the temperature field associated with laser surface treatment is then presented, followed by a discussion on microstructural changes in steels. Materials other than steels are then considered. Next, the hardness and residual stresses that result during laser surface treatment are discussed. Finally, the steps involved, and advantages and disadvantages of the process are presented.

11.1.1 Important Criteria

There are two criteria that are essential for a successful laser surface treatment operation:

1. For steels, the region to be hardened needs to be heated well into the austenite temperature range and be maintained in that temperature range long enough for carbon diffusion to occur.
2. There must be adequate mass in contact with the region to be hardened to permit self-quenching by conduction into the bulk material.

Table 11.1 Typical laser heat treatment process parameters.

Parameter	Value	Units
Power	0.5–9	kW
Defocused beam diameter	2–20	mm
Power densities	1–100	kW/cm^2
Scanning velocities	5–400	mm/s
Hardened depths	0.2–2.5	mm

11.1.2 Key Process Parameters

The principal processing parameters that affect laser surface heat treatment are

1. Beam power, beam size, scanning speed, and shielding gas.
2. Beam mode.
3. Beam absorption by the workpiece.
4. Workpiece material properties and initial microstructure.

The process outputs include the hardness, depth of the hardened region, microstructure, and residual stresses induced.

These key process parameters are discussed in greater detail in the following Sections 11.1.2.1–11.1.2.4.

11.1.2.1 Beam Power, Size, Speed, and Shielding Gas Typical process parameters associated with laser heat treatment are listed in Table 11.1.

A reduction in the scan rate reduces the surface hardness, but increases the depth of the hardened zone (Figure 11.2). This is because reducing the scanning velocity increases the surface temperature and the reaction time for austenitization, resulting in a larger average austenite grain size. Shielding gas (usually argon or helium) may be used to protect the surface from oxidation with flow rates similar to those used during laser welding, about 10 l/min. Due to its higher thermal conductivity, helium results in higher hardness, all other parameters being constant.

Due to the relatively small size of the laser beam, it is often necessary to scan the surface a number of times, with the beam being shifted a specific amount for each scan. To ensure that the entire area of interest is treated, there has to be some amount of overlap between the scans. However, the overlap may anneal portions of the previously hardened structure. Whereas such annealing may not significantly affect the wear properties of a material, it could have significant impact on the fatigue properties, with the softened regions providing preferential sites of stress relaxation at the surface.

11.1.2.2 Beam Mode In contrast to laser cutting, low order beam modes (for example, Gaussian mode) are not particularly suitable for laser surface treatment because of the lack of uniformity in intensity distribution. A relatively wide and uniformly distributed beam is required to obtain uniform surface heating over a wide area and thus avoid localized surface melting. This is done using beam-shaping techniques.

Among the methods commonly available for shaping a laser beam into a more uniform distribution are the following:

1. Beam defocusing using a lens.
2. Optical integration.
3. Beam rastering or scanning.

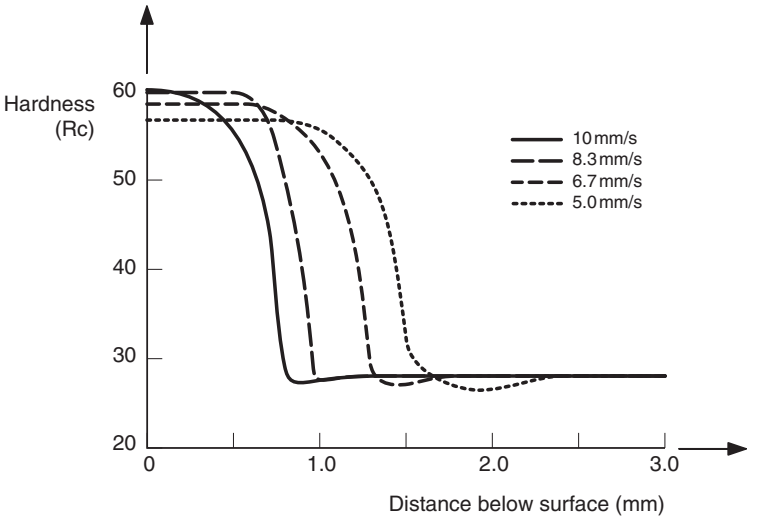


Figure 11.2 Hardness variation for different scanning velocities for AISI 4340 steel tempered at 649 °C for two hours and heat treated with a 1.8 kW CO_2 laser. Source: From Shiue and Chen (1992)/with permission of Springer Nature.

Beam Defocusing This approach is based on normal focusing using a lens. The beam shape is similar to that of the original beam, and the size is changed by varying the distance between the focusing lens and the workpiece or by using an appropriate lens. Thus, the focal point is positioned above the workpiece surface. A focusing system with a very high *F* number or long focal length may be more effective for this method. The F number is the ratio of the focal length of a lens to its diameter. Since beam defocusing does not necessarily produce a uniform beam distribution, the resulting penetration depth distribution across the heat treated track is also non-uniform. Only about 20–30% of the track width may be of uniform penetration depth, depending on the original beam mode. Since the focusing lens is normally part of the beam delivery system, this method is the easiest to implement. It requires no additional effort or expense.

Optical Integration Optical integration involves segmenting the beam into a large number of portions and superposing the individual segments on the same focal plane using a beam integrator. A beam integrator has a number of small mirror segments mounted on a base plate, that is, a multifaceted mirror (Figure 11.3a). Contiguous mirror segments are positioned to be in close contact with each other to minimize beam loss between the segments. The outcome of optical integration is illustrated in Figure 11.3b where a Gaussian beam is segmented into four portions of equal size, and then superposed at the focal plane. Two of the segmented beams are of one orientation, while the other two are of opposite orientation, resulting in near-perfect intensity averaging. In normal cases where the beam is not necessarily symmetrical, as many mirror segments as possible are used.

One drawback of beam integrators is the interference effects that are induced in the integrated beam as a result of the mirror edges. Each edge results in diffraction effects and their superposition produces interference fringes. The higher the Fresnel number, the lower the diffraction losses, and thus the more uniform the integrated beam. The Fresnel number typically ranges between 6 and 12 (low) for infrared wavelength beams. A high Fresnel number would be about 100.

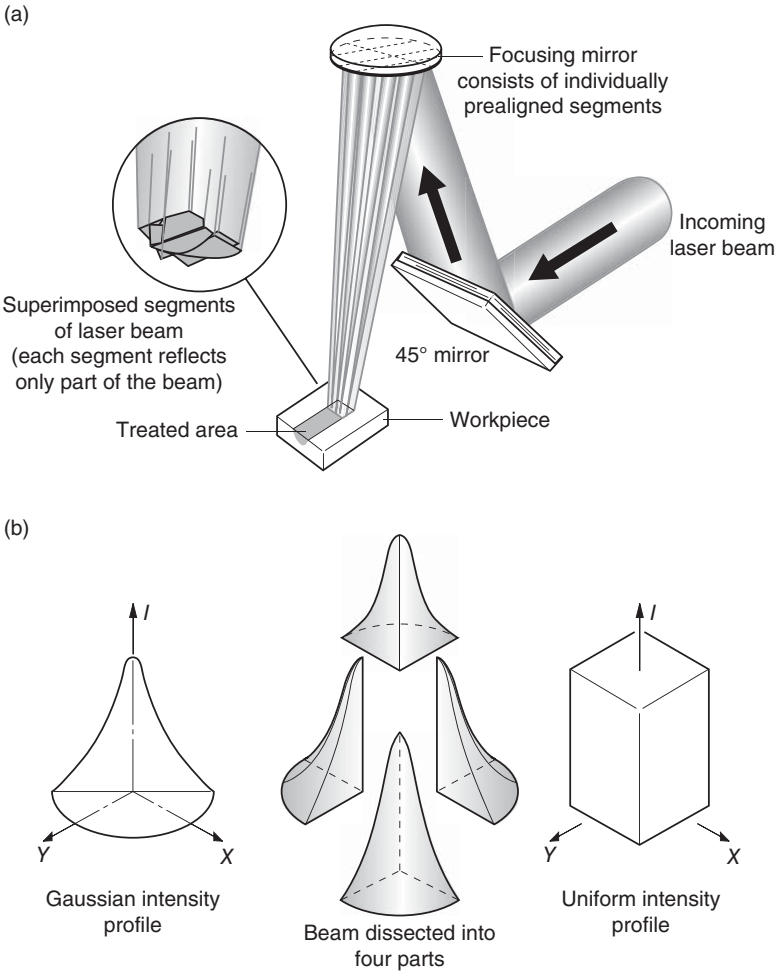


Figure 11.3 Optical integration. Source: Adapted from Mazumder (1983)/Pearson.

For a resonator consisting of mirrors of side $2a \times 2a$, with a spacing L between the mirrors (Figure 11.4), the Fresnel number, N, is given by

$$N = \frac{a^2}{L\lambda} \tag{11.1}$$

The angular spread that results from diffraction of a plane electromagnetic wave of transverse dimensions $2a \times 2a$ is given by the semiangle, θ_d:

$$\theta_d = \frac{\lambda}{2a}$$

However, the semiangle subtended by each of the mirrors at the center of the other is

$$\theta_c = \frac{a}{L}$$

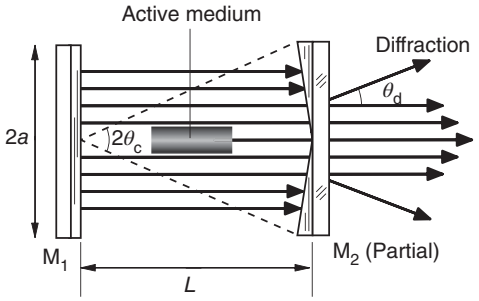


Figure 11.4 Schematic of the resonator geometry and diffraction of the output beam.

Thus, the Fresnel number may be expressed as

$$N = \frac{\theta_c}{2\theta_d} \tag{11.2}$$

What this means is that, for the diffraction losses to be high, the Fresnel number has to be low.

Optical integration generally has the most uniform beam distribution, resulting in about 60–70% of the track width being of uniform penetration depth.

The optics for integrating an output laser beam generally consists of three main components:

1. A flat mirror for beam bending.
2. A spherical convex mirror for expanding the laser beam to fully illuminate the face of the beam integrator for efficient beam integration.
3. A spherical concave beam integrating mirror, which segments the beam into several small beams.

Installation of the optical integration system is easy, but requires more effort than beam defocusing.

Beam Rastering or Scanning A uniformly distributed beam can also be achieved by rastering a finely focused beam to cover a wide area (Figure 11.5). The process involves vibrating two mirrors to get the beam to move back and forth at a high frequency to create the required pattern. Like the optical integration system, installation of the beam rastering system is also easy, but again requires more effort than beam defocusing.

Summary of Beam-Shaping Methods The characteristics of the three types of beam-shaping devices are summarized in Table 11.2.

Figure 11.6a–c illustrate three common forms of beam shape that are used. The circular or elliptical pattern Gaussian beam can be broadened by defocussing; the uniform pattern rectangular beam can be obtained by optical integration; and the rectangular pattern Gaussian beam is achieved by rastering or oscillating a Gaussian beam at a frequency of several hundred Hertz. The corresponding geometry of the hardened and heat-affected zones are also shown in the figures. It is obvious from the figures that the softened heat-affected zone at the part surface is significantly wider for the circular Gaussian beam. The rectangular patterns reduce the width of this zone. However, the depth of the softened region is smaller for the circular Gaussian beam.

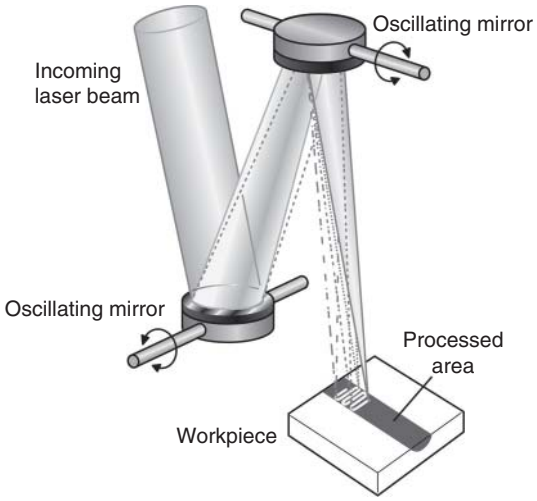


Figure 11.5 Beam rastering. Source: Adapted from Mazumder (1983).

Table 11.2 Comparison of the three beam-shaping devices.

	Focusing lens	Integration	Beam rastering
Power loss	Lowest (1–2%)	Moderate (7–8%)	Low (4–6%)
Energy uniformity	Good	Better	Variable
Cost	Moderate	High	High
Penetration	High	Moderate	Variable
Penetration uniformity	Uneven (20–30%)	Even (60–70%)	Even
Limitations	Fragility of lens	Device length, fragility	Scan rate might limit travel speed
Installation	Very easy	Easy	Easy
Applications	All applications	Heat treating, cladding, alloying	Heat treating, cladding, alloying

11.1.2.3 Beam Absorption The heat treatment process is strongly influenced by

1. Absorption of the beam, which depends on the beam wavelength.
2. Angle of incidence.
3. Material properties or core microstructure.
4. Workpiece surface conditions (finish, coating, and so on).

The high reflectivity of the laser beam when used on metals is a major disadvantage of laser surface treatment. This is especially true of the CO_2 laser beam when incident on metals like copper, gold, and silver. Some of the common techniques that are used to improve absorptivity of reflective materials for effective processing include the following:

1. Applying a thin layer of a highly absorbing coating to the material surface.
2. Using a linearly polarized laser beam on an uncoated surface.
3. Preheating the workpiece, since absorptivity increases with temperature.

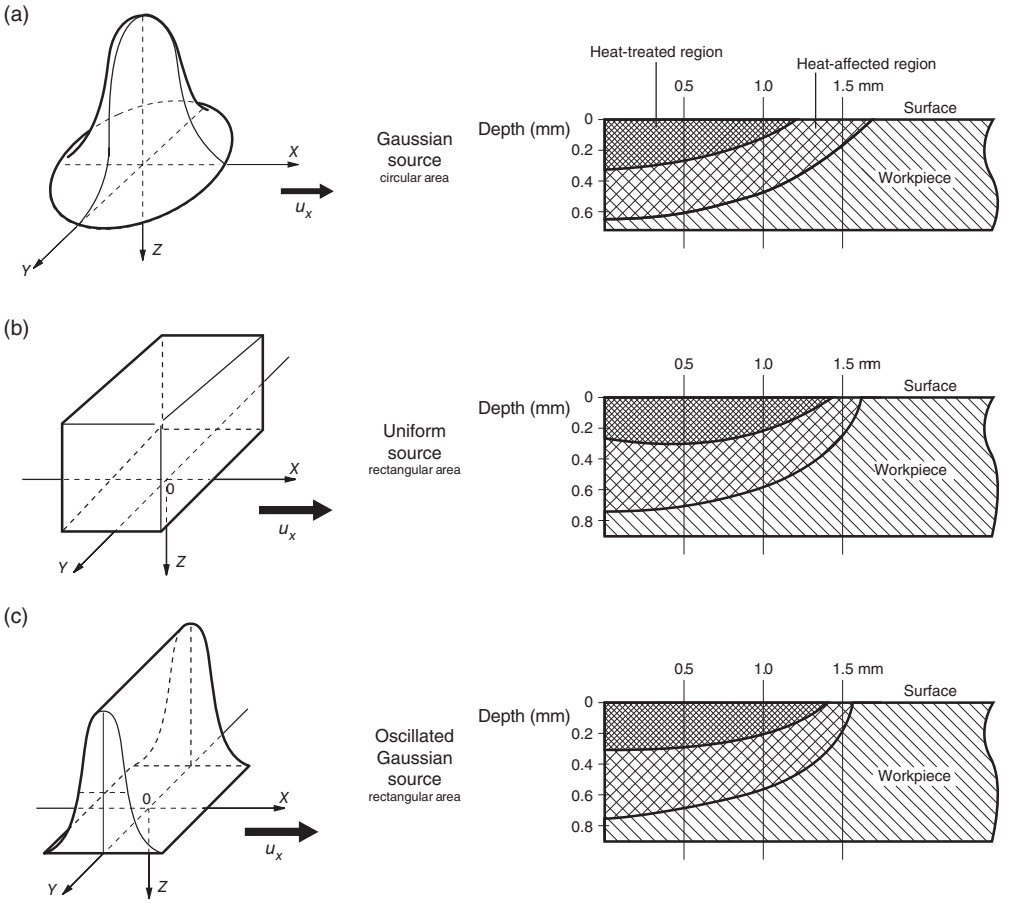


Figure 11.6 Common beam shapes used in laser surface treatment and their corresponding hardened layer geometry. (a) Circular or elliptical pattern Gaussian beam. (b) rectangular beam. (c) Rectangular pattern Gaussian beam. Source: From Seaman (1986).

4. Changing the composition of the material surface by oxidation before processing. However, this approach tends to be more expensive and may not be acceptable for esthetic reasons. Thus, it will not be further discussed.

In the following sections, the first three methods of enhancing absorption are further discussed.

Surface Coating When a surface coating is used, it must have the following characteristics:

1. High thermal stability.
2. Good adhesion to work surface.
3. Chemically passive to workpiece.
4. Easy to remove after treatment.

Examples of coatings normally used are the following:

1. A dispersion of carbon black in alcohol with a binder (colloidal graphite).

2. Chemically deposited copper selenide layer (about 2 μm thick).
3. Manganese, zinc, or iron phosphate (about 2–100 μm thick).
4. Black paint (about 10–20 μm thick).

For these, the fraction of the incident energy that is absorbed by the surface when coated is estimated to be in the range of 60–80% for a CO_2 laser, and this can vary during interaction of the laser beam with the absorptive coating. Graphite is easily applied and is inert, but is relatively more expensive. An excessive coating tends to reduce penetration, as some of the energy absorbed may not be easily transmitted to the base material.

In addition to increasing beam absorption, a carefully applied coating ensures reproducibility of the process. Essentially, all the power that enters the material is absorbed at the surface, and the depth and width of the transformed region depends on the resulting temperature distribution.

Certain coatings, especially the phosphates, have the disadvantage that during processing, they may impregnate along grain boundaries of the base material, Figure 11.7, to form low melting compounds where failure can be initiated. Surface coatings also have the disadvantage that they add to the cost of the process and may prove to be an environmental hazard (especially paints) when they are vaporized during heating. Also, the absorptivity is influenced by variations in thickness or adhesion properties of a surface coating.

Linearly Polarized Beam As discussed in Chapter 3, when an unpolarized beam is incident on a surface at the polarizing angle, the reflected beam is linearly or plane polarized. This is because the component of the electric vector parallel to the plane of incidence vanishes, while the perpendicular component is reflected (Figure 3.2).

When the plane polarized beam is incident on a surface, depending on its orientation, it can have components parallel (p-polarized) and perpendicular (s-polarized) to the plane of incidence. The parallel component is efficiently absorbed at high angles of incidence while the perpendicular component is highly reflected. Thus by using an angle of incidence greater than, say 60°, for a linearly polarized beam, the parallel component (p-polarized beam) will be efficiently absorbed by the surface. The depth of hardening will therefore be greater if a p-polarized beam is used at an incidence angle that corresponds with the angle of maximum absorptivity (Figure 11.8). The drawback of this approach, however, is the significant power loss incurred in the conversion of an unpolarized laser beam to a linearly polarized beam.

Preheating As indicated in Section 8.1.6, the absorptivity of a material increases with an increase in its temperature, for a given wavelength (Figure 8.6). Thus one effective way for enhancing absorption

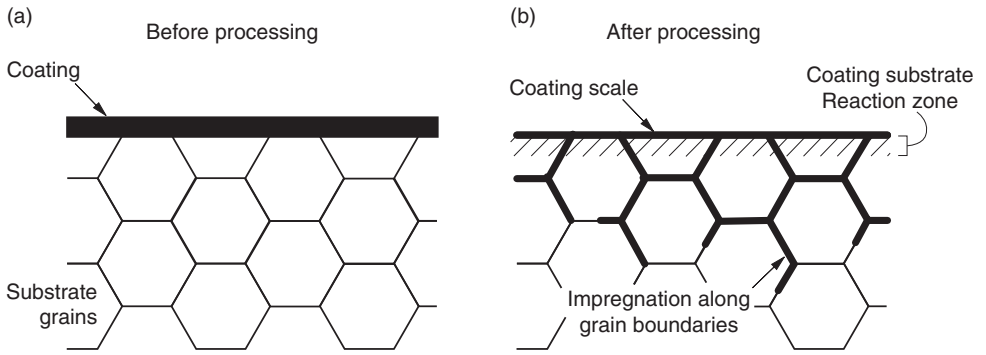


Figure 11.7 Grain boundary impregnation by coatings. Source: Gnanamuthu and Shankar (1985)/SPIE.

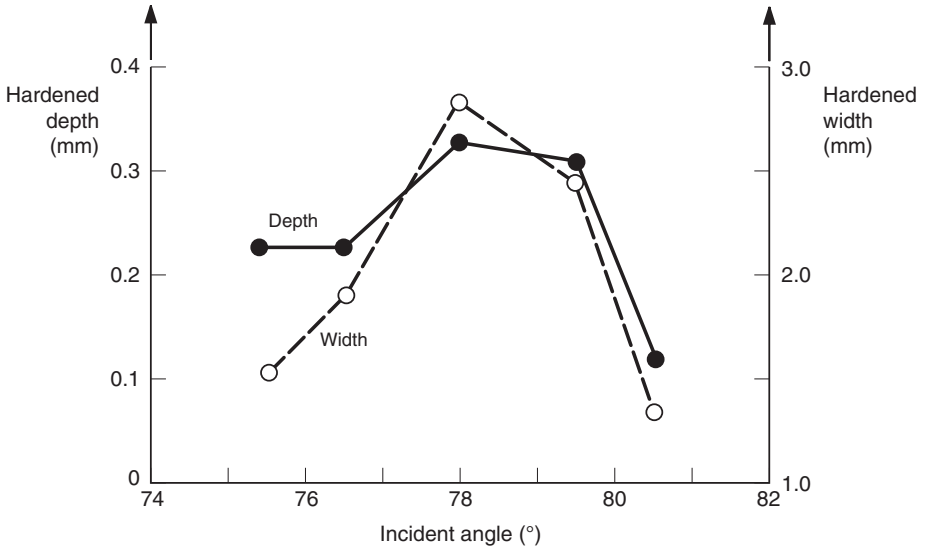


Figure 11.8 Variation of depth of hardening with angle of incidence of a p-polarized beam. Source: From *ICALEO 1991 Proceedings*, Vol. 74, Copyright 1992, Laser Institute of America. All rights reserved.

of the incident laser is to preheat the workpiece before the heat treatment process. The drawback with this approach, however, is that preheating introduces an additional step into the process. Furthermore, the preheating may affect the microstructure of the surrounding material.

11.1.2.4 Initial Workpiece Microstructure The initial microstructure (for example, size of ferrite grains and pearlite colonies) of the workpiece, prior to laser heat treatment, influences the depth of hardening that can be achieved. This is because the distance that the carbon atoms have to diffuse affects the homogenization of austenite, and thus the rate of structural transformation. When the initial or core microstructure is such that the carbon diffusion distance is very small (fine microstructure with a smaller grain size), deep surface-hardened casings are obtained. Thus finer and more evenly distributed carbide dispersoids in a quenched and tempered structure (martensitic or bainitic microstructure prior to laser heat treatment) readily dissolve at lower temperatures within the available exposure time, and thus produce a better hardening effect. On the other hand, when the initial microstructure is coarse (for example, spheroidized), then the carbon diffusion distance needed for the homogenization of austentite is long. This results in less uniform carbon distribution in the austentite, and thus a thin surface-hardened casing (Figure 11.9). The transition from the full martensite hardness to the core hardness is also found to be more gradual in a coarse structure than in a fine structure.

However, for steels with the same carbon content, both fine and coarse structured samples will have the same surface hardness since the extremely high surface temperature results in homogeneous austenitization in the surface region for both structures. Furthermore, since the thermal properties of steels do not change significantly with carbon and alloy content, the same cooling rates will be obtained for the same heat treating conditions. Alloy content may increase the hardenability of the steel, enabling it to form more martensite for a given cooling rate. Hardenability is a measure of the amount of martensite that forms in the microstructure upon cooling. This is affected by both the alloying elements and carbon content. However, since the hardness of a steel is determined primarily

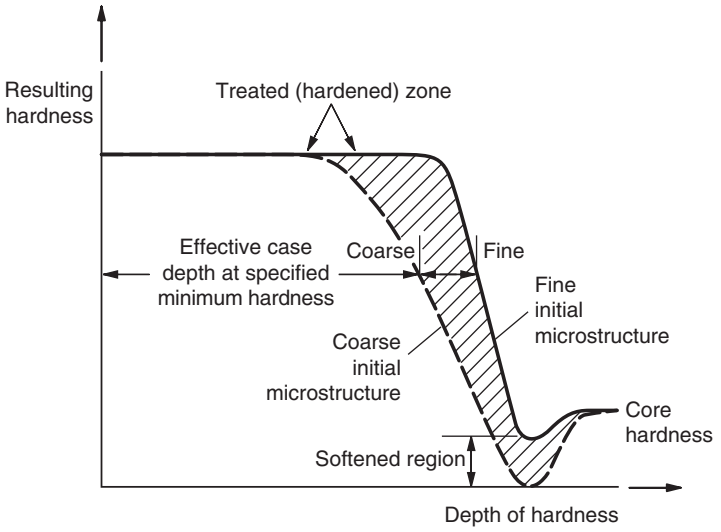


Figure 11.9 Effect of initial microstructure on laser surface treatment.

by the carbon content, an alloy and a plain carbon steel of the same carbon content will have the same hardness of martensite that is formed, but not necessarily the same overall hardness.

11.1.3 Temperature Field

Analysis of heat flow during laser surface treatment can be made using methods that are discussed in Section 4.2. Thus, only a brief reference will be made in this section. Since absorption of the laser beam in a solid metal is confined to a thin surface layer, the z dependence of the absorption can be neglected, and the absorption considered to be two dimensional. Considering a Gaussian distributed heat source moving in a semi-infinite plate in the x-direction, the temperature distribution in the y–z plane (Figure 4.1) can be expressed as

$$T - T_0 = \frac{(1-R)q}{2\pi u_x k[t(t+t_0)]^{1/2}} \times \exp\left\{-\left[\frac{(z+z_0)^2}{4\kappa t} + \frac{y^2}{4\kappa(t+t_0)}\right]\right\} \tag{11.3}$$

where t is the interaction time, $t_0 = \frac{w^2}{4\kappa}$ is the time required for heat to diffuse a distance equal to the beam radius, and z_0 is the distance over which heat can diffuse during the beam interaction time, $\frac{w}{u_x}$. It is a characteristic length that limits the surface temperature to a finite value and is obtained by matching solutions at $z = z_0$ to known solutions for the peak surface temperature.

11.1.4 Microstructural Changes in Steels

Steels are extensively discussed in this section because they are still among the most widely used metal alloys today. Some metals undergo polymorphic transformation (that is, change their crystal structure) when heated above a certain temperature. The change in crystal structure also changes their properties. An example is iron which changes from α-iron with the BCC structure at room temperature to γ-iron with the FCC structure when heated above 910 °C (Appendix 11.B). The solubility

of carbon in γ-iron is greater than in α-iron, and this makes steel subject to significant changes in property when heat treated.

The overall transformation process during heat treatment of steels occurs in three stages:

1. Diffusion-controlled transformation of pearlite to austenite (pearlite dissolution).
2. Diffusion-controlled homogenization of carbon in austenite.
3. Diffusionless or displacive transformation of austenite to martensite.

These three stages are discussed in further detail in the following Sections 11.1.4.1–11.1.4.3.

11.1.4.1 Pearlite Dissolution Pearlite dissolution involves the transformation of the pearlite colonies to austenite as the material is heated. For plain carbon steels, when the material is heated above the eutectoid (A_1) temperature, the pearlite colonies transform to austenite. For pre-eutectoid steels, as the temperature increases, the size of the austenite regions increase by outward diffusion of carbon from the high-carbon (0.8% C) pearlite colonies into the ferrite region. Let the spacing of the pearlite plates in a colony be l_p, Figure 11.10. Then lateral diffusion of carbon over a distance l_p might be considered sufficient to convert the colony to austenite. For an isothermal process, the time, t, required for such diffusion is given by the equation:

$$l_p{}^2 = 2D_a t \tag{11.4}$$

where D_a is the diffusion coefficient for carbon.

However, in the case of heat treatment where the temperature is time-dependent, the term $D_a t$ is replaced by

$$D_a t = \int_0^\infty D_0 \exp\left(-\frac{Q_a}{R_g T(t)}\right) dt \approx D_0 c_d \tau_h \exp\left(-\frac{Q_a}{R_g T_p}\right) \tag{11.5}$$

where c_d is a constant, D_0 is a diffusion constant, Q_a is activation energy for the transformation, R_g is the gas constant, T_p is the peak temperature experienced during heating, and τ_h is the thermal time constant.

Thus,

$$l_p{}^2 \approx 2D_0 c_d \tau_h \exp\left(-\frac{Q_a}{R_g T_p}\right) \tag{11.6}$$

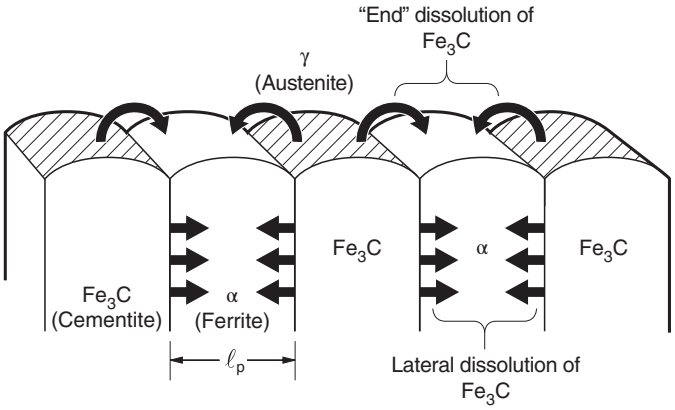


Figure 11.10 Transformation of pearlite to austenite. Source: Adapted from Ashby and Easterling (1984).

For a given pearlite plate spacing, l_p, Eq. (11.6) enables the peak temperature that is necessary for the transformation to take place to be estimated. The constant c_d is approximated by

$$c_d = 3 \times \sqrt{\frac{R_g T_p}{Q_a}} \tag{11.7}$$

To determine the processing conditions that will give the peak temperature, T_p, and thermal time constant τ_h, the following equations may be used:

$$T_{p1} = T_0 + \left(\frac{2}{e}\right)^{1/2} \frac{(1-R)q}{\pi \rho c_p w u_x (z + z_{01})} \qquad t \ll t_0 \tag{11.8}$$

$$\tau_{h1} = \frac{1}{\pi^2 k e \rho c_p} \left[\frac{(1-R)q}{w u_x (T_p - T_0)}\right]^2 \qquad t \ll t_0 \tag{11.9}$$

$$T_{p2} = T_0 + \frac{2(1-R)q}{\pi e \rho c_p u_x (z + z_{02})^2} \qquad t \gg t_0 \tag{11.10}$$

$$\tau_{h2} = \frac{(1-R)q}{2\pi k e u_x (T_p - T_0)} \qquad t \gg t_0 \tag{11.11}$$

where e is the natural exponent, ${z_{01}}^2 = \frac{\pi \kappa w}{2 e u_x}$, and ${z_{02}}^2 = \frac{w}{e}\left(\frac{\pi \kappa w}{u_x}\right)^{1/2}$

Example 11.1 The surface of an AISI 1045 steel cam is to be heat treated using a 2.5 kW CO_2 laser to enhance the wear resistance. The defocused beam diameter is 5 mm, and the scan rate is set at 10 mm/s. Determine the spacing of the pearlite plates that will enable the pearlite colonies to be fully transformed to austenite under these conditions.

Assume:

1. The temperature is time-dependent.
2. The diffusion constant, D_0, for the diffusion of carbon in α-Fe (ferrite) is $0.020\,\text{cm}^2/\text{s} = 2 \times 10^{-6}\,\text{m}^2/\text{s} = 2\,\text{mm}^2/\text{s} = 2 \times 10^6\,\mu\text{m}^2/\text{s}$ (Shewmon 1983, Table 2-1, p. 64).
3. The activation energy for diffusion, Q_a, is 20.1 kcal/mol = 84.15 kJ/mol (Shewmon 1983, Table 2-1, p. 64).
4. The reflection coefficient of the surface is 80%.
5. Ambient temperature is 25 °C.

Solution:

The peak temperature and thermal time constant that will be obtained for the given conditions need to be determined. However, without additional information, we assume that t is indefinite, since we need to use Eq. (11.5). Therefore to calculate T_p and τ_h, we use Eqs. (11.10) and (11.11) on the basis that $t \gg t_0$:

$$T_{p2} = T_0 + \frac{2(1-R)q}{\pi e \rho c_p u_x (z + z_{02})^2} \qquad t \gg t_0 \tag{11.10}$$

$$\tau_{h2} = \frac{(1-R)q}{2\pi k e u_x (T_p - T_0)} \qquad t \gg t_0 \tag{11.11}$$

From Appendices 4.D and 4.E, the properties of steel are approximated as

Average density, $\rho = 7870\,\text{kg/m}^3$
Average specific heat, $c_\text{p} = 452\,\text{J/kg K}$
Thermal conductivity, $k = 0.073\,\text{W/mm K}$.

The thermal diffusivity, κ, of the material is

$$\kappa = \frac{k}{\rho c_\text{p}} = \frac{0.073}{7.87 \times 10^{-6} \times 452} = 20.52\,\text{mm}^2/\text{s}$$

Also,

$${z_{02}}^2 = \frac{w}{e}\left(\frac{\pi\kappa w}{u_x}\right)^{1/2} = \frac{2.5}{2.71828}\left(\frac{\pi \times 20.52 \times 2.5}{10}\right)^{1/2} = 3.69\,\text{mm}^2$$

$$\Rightarrow z_{02} = 1.92\,\text{mm}$$

Since the peak temperature will occur on the surface of the workpiece, $z = 0$. Thus, we have

$$T_{\text{p2}} = T_0 + \frac{2(1-R)q}{\pi e \rho c_\text{p} u_x (z + z_{02})^2} = 25 + \frac{2 \times (1 - 0.8) \times 2500}{\pi \times 2.71828 \times 7.87 \times 10^{-6} \times 452 \times 10 \times (0 + 1.92)^2}$$
$$= 918\,^\circ\text{C}$$

Therefore,

$$\tau_{\text{h2}} = \frac{(1-R)q}{2\pi k e u_x (T_p - T_0)} = \frac{(1 - 0.8) \times 2500}{2 \times \pi \times 0.073 \times 2.71828 \times 10 \times (918 - 25)} = 0.045\,\text{s} = 45\,\text{ms}$$

The gas constant $R_\text{g} = 1.98\,\text{cal/mol K}$. Thus,

$$c_\text{d} = 3 \times \sqrt{\frac{R_\text{g} T_\text{p}}{Q_\text{a}}} = 3 \times \sqrt{\frac{1.98 \times (918 + 273)}{20.1 \times 10^3}} = 1.027$$

Now the pearlite plate spacing is obtained from Eq. (11.6) as

$$l_\text{p}^2 \approx 2 D_0 c_\text{d} \tau_\text{h} \exp\left(-\frac{Q_\text{a}}{R_\text{g} T_\text{p}}\right) = 2 \times 2 \times 10^{-6} \times 1.027 \times 0.045 \times \exp\left(-\frac{20.1 \times 10^3}{1.98 \times (918 + 273)}\right)$$
$$= 0.367 \times 10^{-10}\,\text{m}^2$$

or

$$l_\text{p} = 0.61 \times 10^{-5}\,\text{m} = 6.1\,\mu\text{m}.$$

11.1.4.2 Austenite Homogenization Austenite homogenization follows the transformation from pearlite to austenite and is associated with the diffusion of carbon from the high carbon austenite region (which transformed from pearlite) to the low carbon austenite region (which transformed from ferrite, Figure 11.11). The extent of homogenization depends on temperature and time.

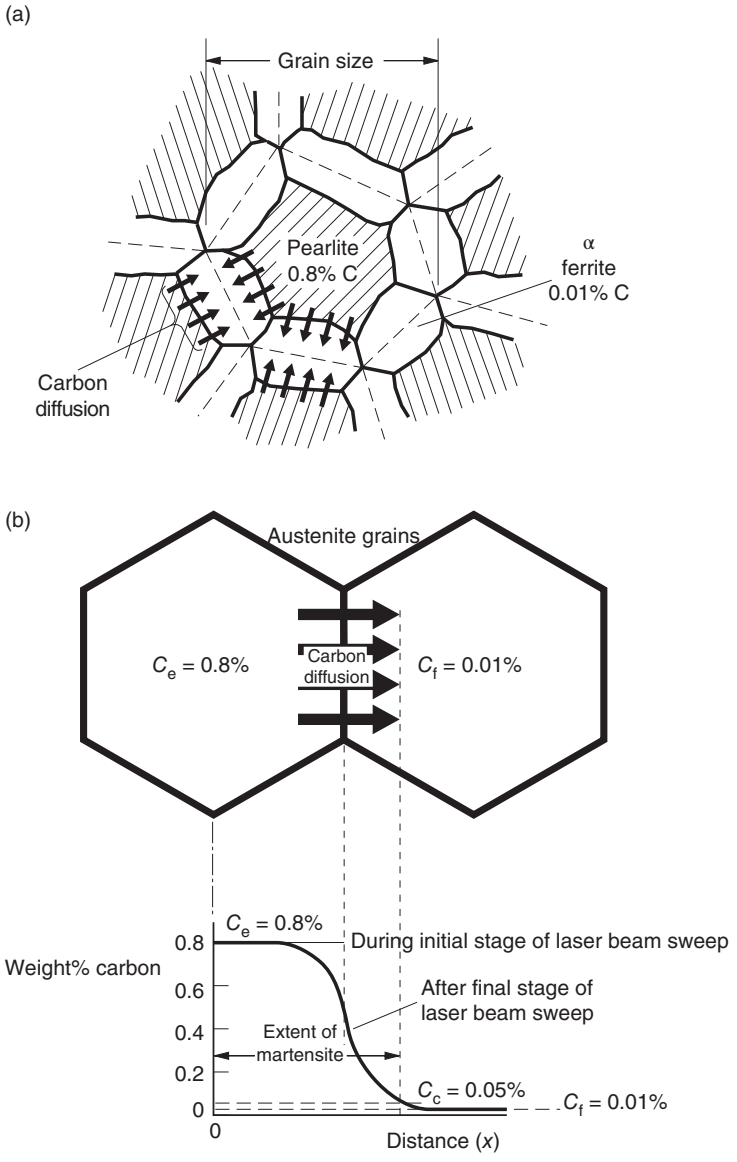


Figure 11.11 Homogenization of austenite. (a) Diffusion of carbon from pearlite to ferrite. (b) Variation of carbon content across the pearlite–ferrite interface. Source: Adapted from Ashby and Easterling (1984).

The variation of the carbon content with time as homogenization proceeds is obtained from the diffusion equation as

$$C_a(x,t) = \frac{1}{2}(C_e + C_f) - \frac{1}{2}(C_e - C_f)\text{erf}\left[\frac{x}{2\sqrt{D_a t}}\right] \tag{11.12}$$

where C_a is the carbon content in wt%, $C_e = 0.8\%$ is the pearlite carbon content, C_f is the carbon content of the ferrite, and x is the distance over which a carbon content of C_a is achieved in time t.

Now if C_c (Figure 11.11b) is the critical carbon fraction above which transformation to martensite occurs, then for $C_f \ll C_e$, the boundary of the region in which $C_a \geq C_c$ is approximated by

$$x = \frac{2}{\sqrt{\pi}} \ln\left(\frac{C_e}{2C_c}\right)\sqrt{D_a t} \tag{11.13}$$

Equation (11.13) defines the size of the zone over which martensite can be formed after the specimen is heated for a given time, t.

Example 11.2 Estimate the average size of the region where martensite can be formed, for the conditions in Example 11.1.

Solution:
Using Eq. (11.13), we have

$$x = \frac{2}{\sqrt{\pi}} \ln\left(\frac{C_e}{2C_c}\right)\sqrt{D_a t}$$

C_c is obtained from Figure 11.11 as $C_c = 0.05\%$. Thus,

$$x = \frac{2}{\sqrt{\pi}} \ln\left(\frac{C_e}{2C_c}\right)\sqrt{D_0 c_d \tau_h \exp\left(-\frac{Q_a}{R_g T_p}\right)}$$

$$x = \frac{2}{\sqrt{\pi}} \ln\left(\frac{0.8}{2\times 0.05}\right)\sqrt{2\times 10^{-6}\times 1.027\times 0.045\times \exp\left(-\frac{20.1\times 10^3}{1.98\times(918+273)}\right)}$$

$$= 1.01\times 10^{-5}\ \text{m} = 10.1\ \mu\text{m}$$

11.1.4.3 Transformation to Martensite Martensite is normally formed when austenite is cooled at a high enough rate. It is extremely hard and brittle. The transformation from austenite to martensite is diffusionless and takes place almost instantaneously.

The volume fraction f of martensite that is formed over a period t is given by

$$f = f_m - (f_m - f_i)\exp\left[-\frac{12 f_i^{2/3}}{g_s\sqrt{\pi}} \ln\left(\frac{C_e}{2C_c}\right)\sqrt{D_a t}\right] \tag{11.14}$$

where $f_i = \frac{C_a - C_f}{0.8 - C_f} \approx \frac{C_a}{0.8}$ is the volume fraction initially occupied by the pearlite colonies, which is also the minimum subsequent volume fraction of martensite,

$$f_m = \begin{cases} 0 & \text{if} \quad T_p < A_1 \\ f_i + (1 - f_i)\dfrac{T_p - A_1}{A_3 - A_1} & \text{if} \quad A_1 < T_p < A_3 \\ 1 & \text{if} \quad T_p > A_3 \end{cases}$$

g_s is the average austenite grain size; f_m is the maximum volume fraction permitted by the phase diagram; A_1 is the eutectoid temperature (Appendix 11.B); and $A_3 = 1183 - 416C_a + 228C_a^2$ K is the lower temperature boundary of the austenite zone in the iron–carbon diagram. For the situation where the temperature is time-dependent, the term $D_a t$ in Eqs. (11.12)–(11.14) is replaced by $D_0 c_d \tau_h \exp\left(-\frac{Q_a}{R_g T_p}\right)$.

Example 11.3 Determine the volume fraction of martensite that can be formed, for the conditions in Example 11.1. Assume that the average austenite grain size is 200 μm.

Solution:
The volume fraction f can be obtained from Eq. (11.14) as

$$f = f_m - (f_m - f_i)\exp\left[-\frac{12 f_i^{2/3}}{g_s\sqrt{\pi}}\ln\left(\frac{C_e}{2C_c}\right)\sqrt{D_a t}\right]$$

Now

$$f_m = \begin{cases} 0 & \text{if} \quad T_p < A_1 \\ f_i + (1 - f_i)\dfrac{T_p - A_1}{A_3 - A_1} & \text{if} \quad A_1 < T_p < A_3 \\ 1 & \text{if} \quad T_p > A_3 \end{cases}$$

But

$$f_i = \frac{C_a - C_f}{0.8 - C_f} \approx \frac{C_a}{0.8} = \frac{0.45}{0.8} = 0.5625 \text{ for 1045 steel since } C_a \text{ is then 0.45.}$$

Also, $A_1 = 723\ ^\circ\text{C}$, and the A_3 temperature is estimated to be $A_3 = 769\ ^\circ\text{C}$.
Thus, since $T_p = 918\ ^\circ\text{C}$ from Example 11.1, $f_m = 1$. Therefore, we have

$$f = f_m - (f_m - f_i)\exp\left[-\frac{12 f_i^{2/3}}{g_s\sqrt{\pi}}\ln\left(\frac{C_e}{2C_c}\right)\sqrt{D_0 c_d \tau_h \exp\left(-\frac{Q_a}{R_g T_p}\right)}\right]$$

$$= 1 - (1 - 0.5625)\exp\left[-\frac{12 \times 0.5625^{2/3}}{200 \times 10^{-6}\sqrt{\pi}}\ln \times\left(\frac{0.8}{2 \times 0.05}\right)\sqrt{2 \times 10^{-6} \times 1.027 \times 0.045 \times \exp\left(-\frac{20.1 \times 10^3}{1.98 \times (918 + 273)}\right)}\right]$$

Thus, the volume fraction, f, of martensite formed is

$$f = 64.4\%.$$

11.1.5 Nonferrous Alloys

Even though our discussion thus far has focused on phase transformation in steels, other materials can also be subjected to laser heat treatment to improve their mechanical properties. The hardness of some aluminum alloys, for example, can be increased by about 100% after laser treatment. The principal process for increasing the strength of nonferrous alloys is precipitation hardening. This requires the existence of a strong supersaturated solid solution at high cooling rates. Thus, it is only possible in alloys in which the solubility of one element in the solid solution increases with increasing temperature (Figure 11.12). There are two steps involved in the precipitation hardening process:

1. Solution treatment.
2. Aging.

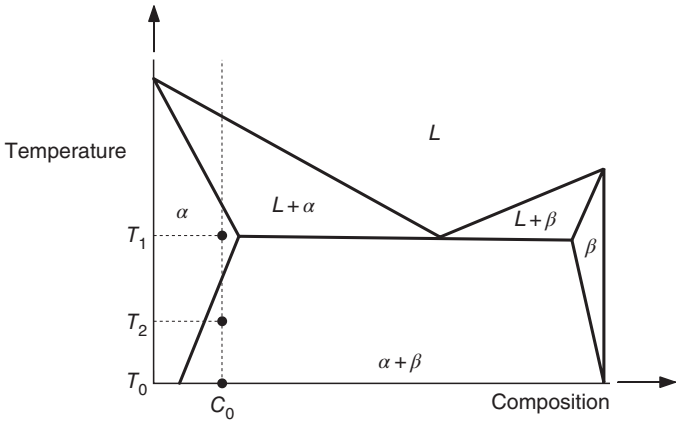


Figure 11.12 Solution treatment.

11.1.5.1 Solution Treatment The material is heated into the portion of the phase diagram where only a solid solution phase exists, say T_1 (Figure 11.12), and allowed to stay until the solid solution phase, α, is formed. It is then quenched to T_2. The rapid cooling prevents the formation of the second phase, β, and the alloy formed at this stage is supersaturated, and initially relatively soft.

11.1.5.2 Aging If the alloy is maintained at T_2 for a while, nuclei of the second phase are formed and begin to grow throughout the material, resulting in second phase particles. The formation of the second phase particles impedes the motion of dislocations. That means more stress is required to move the dislocations through them, and the strength of the material thus increases (Figure 11.13a). The more particles are formed and grow, the greater the strength becomes (stage 1). Thus, a finer distribution of second phase particles produces a stronger material. However, eventually the strength reaches a maximum (stage 2) and a further stay at T_2 only results in continued growth of the particles as smaller particles join together to form larger ones, and the spacing between them increases. The strength then begins to decrease (stage 3). This is overaging.

The lower the aging temperature, T_3, the longer it takes to reach the maximum strength, but the greater is the maximum strength since the second phase is more finely dispersed at the lower temperatures due to the increased nucleation rate (Figure 11.13b,c). If T_3 is low enough, the second phase may not form at all.

Nonferrous alloys may also be hardened by laser treatment if at least one of the following conditions is satisfied:

1. Formation of a characteristic finely-divided structure in the hardening zone.
2. Formation of metastable phases.
3. Dislocation density increase.

11.1.6 Hardness Variation

The variation of hardness and ductility for a surface-hardened workpiece is illustrated in Figure 11.14. The top view (Figure 11.14a) illustrates hardened regions where the hardness is higher and the ductility lower than the original material. As the cross-sectional profile of the surface-hardened region shows (Figure 11.14b), hardness is a maximum at the surface and gradually

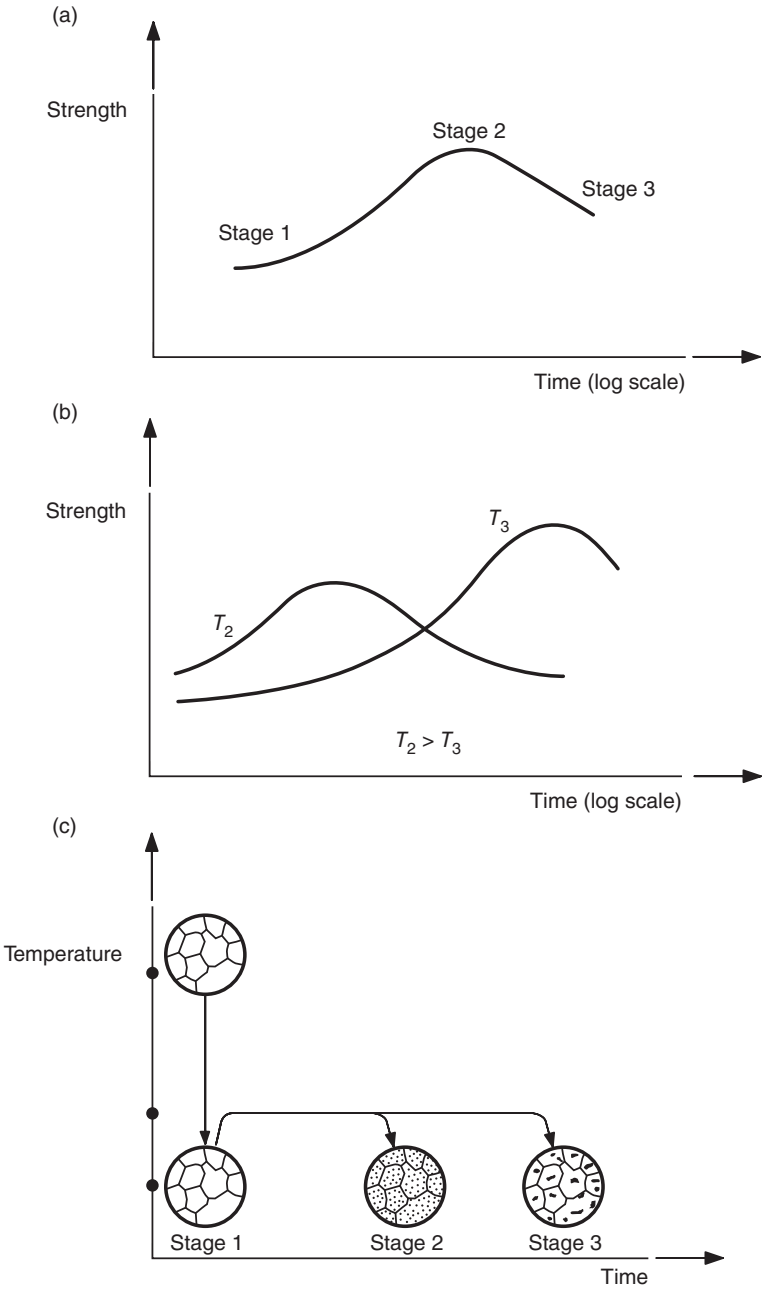


Figure 11.13 Illustration of the aging process. (a) Variation of hardness with time during aging. (b) Temperature dependence of the aging curve. (c) Variation of microstructure with time during aging. Source: From Barrett et al. (1973).

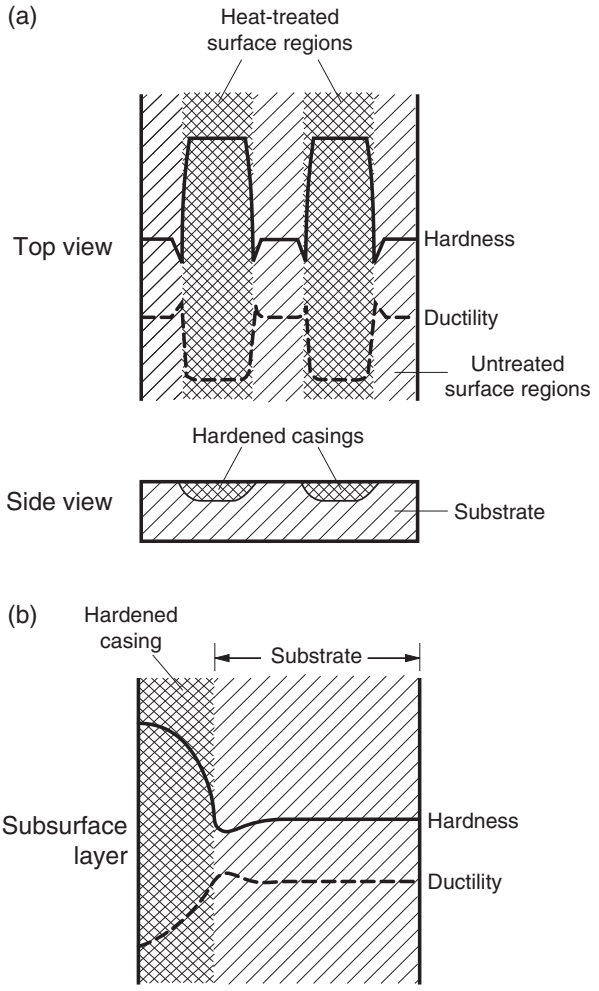


Figure 11.14 Hardness and ductility variation on a surface-hardened workpiece. (a) Hardness distribution on the top surface of the workpiece. (b) Hardness variation with depth. Source: Gnanamuthu and Shankar (1985)/SPIE.

decreases with depth from the surface. This is the trend observed for macrohardness measurements that may be obtained using a Vickers indentor. However, microhardness measurement of individual martensitic regions for pre-eutectoid steels shows the martensite hardness to increase with distance below the surface. This is because the carbon content of the individual martensite formed is highest when the peak temperature experienced is just above the A_1 (eutectoid) temperature of the iron-carbon phase diagram (Appendix 11.B). The A_1 temperature is achieved at a distance farther away from the heat source. At this temperature, the carbon content of the austenite that eventually transforms into martensite is about 0.8%, which is higher than that of the original steel. Also, at higher peak temperatures, carbon is able to diffuse faster outward into the surrounding low carbon matrix. This reduces the local carbon concentration. Now since the hardness of martensite decreases with decreasing carbon content, the hardness of the individual martensitic regions is lower in the area where the peak temperatures are highest. On the contrary, more martensite is formed in the region where the peak temperatures are highest. Thus, the overall hardness in this region is higher

than that in the lower temperature region, since the overall hardness of a material increases with the amount of the harder material.

The drop in hardness observed next to the hardened region is the part of the substrate which is heated to a high enough temperature for grain growth to occur, but where the cooling rate is not high enough to result in martensite formation.

11.1.7 Residual Stresses

Residual stresses are normally induced in the surface layers of a part that is subjected to surface modification. The key mechanisms responsible for the residual stresses are the following:

1. Thermal expansion and contraction associated with the process (Chapter 7).
2. Phase transformation, as occurs during austenite to martensite transformation. The second mechanism is due to an increase in volume as austenite transforms to martensite, resulting in compressive stresses. These compressive residual stresses improve the fatigue life of the part.

The variation of transverse stresses across the laser track for a single scan laser heat treatment is shown in Figure 11.15, indicating tensile transverse stresses within the laser track, with balancing compressive stresses in the heat-affected zone outside the track.

11.1.8 Semiconductors

Polycrystalline silicon that is deposited on a dielectric substrate such as glass is useful for thin film transistors (TFTs), which are used in liquid-crystal display (LCD) technology, and also for thin film solar cells. The solar cells are more efficient when the grain size is larger than the film thickness. The electrical properties of the TFT also depend on the grain structure (size and texture) of the polycrystalline silicon. Better electrical properties are obtained when the silicon has a large grain structure. Such large grain structure can be achieved by crystallizing a layer of amorphous silicon with a laser. Thickness of the amorphous silicon film may be about 150 nm and is normally deposited by chemical vapor deposition.

Pulsed lasers have an advantage in being used for crystallizing the silicon layer, as opposed to furnace hearting, since it is possible to control the beam parameters to avoid cracking or melting the

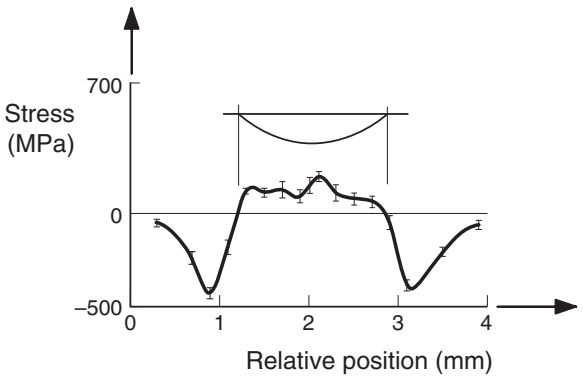


Figure 11.15 Transverse residual stress variation on a surface-hardened workpiece. Source: Van Brussel and De Hosson (1993)/with permission of Elsevier.

glass substrate. The use of excimer lasers is well established, but the low pulse repetition rate of about 1 kHz limits the scanning speed, making the process slow. Nd:YVO_4 lasers with pulse repetition rates up to 100 kHz result in significant improvements in the crystallization rate with similar enhancements in the electrical properties. These are typically laser diode pumped, and the wavelength frequency doubled to 0.532 μm, and typical processing parameters would involve a pulse energy of about 40 μJ and a scanning speed of 10 mm/s.

11.1.9 Polymers

The characteristics of polymer surfaces can be changed with appropriate treatment using a laser. For example, surfaces that repel water can become hydrophilic, as demonstrated by a reduced contact angle, after laser treatment using a linearly polarized KrF or ArF excimer laser. In some cases, electrical conductivity is also improved. This is especially the case with thermostable polymers such as polyimides with high glass transition temperature. The surface texture that is created can also affect characteristics in areas such as tribology, optics, and adhesion.

Nanostructures have also been created on polymer surfaces using pulsed Nd:YAG lasers that have the advantage that they do not require additional polarization. Such structures facilitate the fabrication of fine gratings and quantum devices. Periodic linear structures are produced after a few hundred pulses, when a frequency quadrupled (fourth harmonic) Nd:YAG laser output of wavelength 0.266 μm is introduced on the surface of a polymer such as polyimide. Typical process parameters would be a pulse repetition rate of 10 Hz, pulse duration of 5 ns, and fluence of 3–5 mJ/cm^2, which is below the polymer's ablation threshold. The period, Λ, of the linear structure that results is given by

$$\Lambda = \frac{\lambda}{n - \sin\theta}$$

where λ is the laser wavelength, n is the material refractive index, and θ is the angle of incidence of the laser beam. The linear structure is the result of a periodic electromagnetic field which is created by the interference of a surface scattered radiation and the incident beam. The ripple direction is parallel to the plane of polarization of the laser beam.

A typical width of the lines would be about 100 nm, with a depth of 500 nm. Continued exposure of the polymer surface to the laser pulses will result in conversion of the linear structure to a dot pattern after about 8000 pulses. For some polymers, such as the polystyrene family, the presence of oxygen (processing in air) is necessary for the surface pattern to be created.

11.1.10 Advantages and Disadvantages of Laser Surface Treatment

11.1.10.1 Advantages The primary advantages of laser surface heat treatment include the following:

1. *Rapid Heating and Cooling*: This enables steels of lower hardenability to be heat treated. Thus, it is often suitable for components that are difficult to harden using conventional methods such as flame and induction hardening.
2. The resulting heat-affected zone is minimal, compared to that associated with flame and induction hardening.
3. Minimal distortion due to low heat input. This eliminates post-hardening machining and associated cost.
4. Localized heat input enables only desirable regions to be heat treated. Other regions are thus essentially unaffected. It also minimizes the total heat input to the part. Thus, less energy is used.

5. Ease of processing complex shapes due to the ability to scan the beam over the part.
6. More uniform case hardness.
7. No external quenching is necessary since the process often involves self-quenching.
8. Minimization of fumes and dirt that results from heating and quenching.
9. The short cycle time results in relatively fine grain structure. This enables higher strength and good fatigue resistance to be achieved.

11.1.10.2 Disadvantages

1. Due to the short interaction times, coarse structures (or alloys that require long soaking times for heat treatment) tend to be difficult to laser heat treat. These include coarse pearlite, blocky ferrite, and steels that contain spheroidal carbides and cast irons that consist primarily of graphite without any pearlite. The austenite may not be completely homogenized in such materials, and thus soft spots may occur in the hardened zone.
2. High capital cost of the laser.
3. The need for sufficient mass for self-quenching.

11.2 LASER SURFACE MELTING

Steels with widely dispersed carbide or graphite (mainly low carbon steels) are generally difficult to harden by solid-state transformation. These are more easily hardened by melting and solidification, since the diffusion rate of carbon is much higher in the molten state.

The localized nature of a laser as an energy source makes it effective for melting a thin surface layer of metal to control its microstructure.

Laser surface melting of metals (also known as skin melting or glazing) involves melting of a thin surface layer that subsequently undergoes rapid solidification as a result of self-quenching, resulting in alterations in the local microstructure. The process is similar to autogenous (no filler metal added) conduction mode laser welding. However, in this case, lower power densities are used at higher traverse speeds to ensure that only a thin layer of the substrate is melted, and that the resulting cooling rates are high enough to induce rapid solidification. Thus, the principles of rapid solidification discussed in Section 6.3 apply to this process.

The microstructural changes may be accompanied by changes in properties such as hardness, corrosion resistance, and wear resistance. The surface melting process results in four distinct zones in the material (Figure 5.1):

1. Fusion zone.
2. Zone of partial melting.
3. Heat-affected zone.
4. Base material.

When the process involves multiple scanning, there also exists an overlapping zone when the individual beam scans are made to overlap.

The microstructure of the fusion zone depends on the alloy. For example, for AISI 4340 steel, it may consist of a fine cellular structure containing martensite, while for AISI 4140, it may consist primarily of fine grained martensite with traces of austenite. The microhardness of the fusion zone is generally higher than that of the overlapping zone, whose hardness is in turn higher than that of the

base material. Unlike the case of surface heat treatment (where no melting occurs), the microstructure of both the fusion and overlapping regions is independent of the thermal history of a steel, while that of the HAZ depends on the original matrix. The microstructure of the HAZ may consist of coarse martensite, since it is austenitized during heating, and then quenched.

In the fusion zone, increasing the scanning speed or decreasing the laser power results in a finer microstructure and thus higher microhardness. However, increasing the laser power increases the melt depth. For a single scan treatment, the microhardness is highest at the center of the path and for overlapping treatments reduces towards the overlap zone. Reducing the amount of overlap improves the overall microhardness.

11.3 LASER DIRECT METAL DEPOSITION

Laser deposition is a form of cladding, hardfacing, or coating technique for enhancing the corrosion and/or wear resistance of a part. This process involves using a laser beam to melt a very thin surface of a substrate. The molten substrate mixes with melted clad alloy which is usually applied as powder (Figure 11.16) or wire feedstock to form a metallurgical bond. As in the other cases of laser surface modification, the rapid solidification and/or quench rates associated with the process enable either amorphous or nonequilibrium crystalline phases to be formed. The amorphous phases enable material properties to be obtained that cannot be obtained under equilibrium cooling. Laser cladding applications include engine valve components, boiler firewall, and turbine blades. It also serves as an effective tool for repairing components.

Potential problems that may arise during cladding include the following:

1. Porosity that may result from moisture and the entrapment of shielding gas.
2. Thermal stresses that may cause cracking and are due to the following factors:
   (a) Difference between the melting points of the clad layer and substrate.
   (b) Different coefficients of thermal expansion of the clad layer and substrate.
   (c) Volume changes associated with phase changes that occur as the material cools down.

11.3.1 Processing Parameters

As an illustration, typical processing parameters for cladding of an Mg–Al alloy, for example, by powder feeding are listed in Table 11.3.

For Mg–Al alloys, the use of nitrogen shielding gas may result in the greatest depth of the alloyed region. Addition of helium or argon reduces the melt depth, but increases the hardness level by as much as 100%. Inert gas mixtures without any nitrogen result in the hardest cladding, with minimal porosity or cracking.

The clad tracks that are formed in laser deposition tend to be discontinuous with a high height-to-width (aspect) ratio at relatively low speeds, becoming shallower with a reducing aspect ratio as speed increases (Figure 11.17). Thus at very high speeds, about 30 mm/s, the cladding obtained is relatively thin, and the dilution (Section 11.3.3) increases accordingly.

11.3.2 Methods for Depositing the Material

Material may be deposited in a number of ways. These include the following:

1. Preplaced loose powder bed.
2. Presprayed surfacing powder.
3. Prepositioned chip or rod.

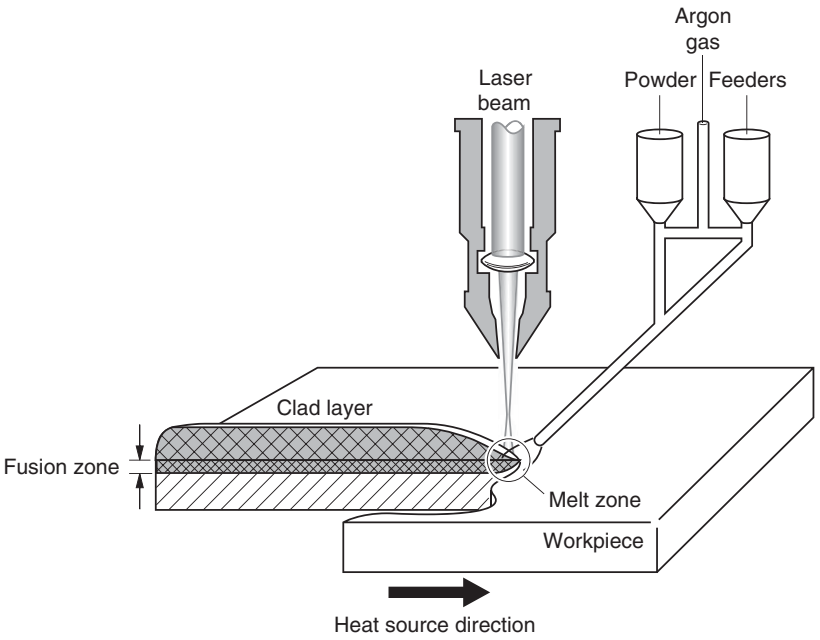


Figure 11.16 Schematic of the cladding process – The powder feed method. Source: Jasim and West (1989)/SPIE.

Table 11.3 Typical process parameters used in laser cladding.

Parameter	Value	Units
Power	1–3	kW
Defocused beam diameter	3–5	mm
Power densities	1–100	kW/cm^2
Scanning velocities	1–60	mm/s
Powder feed rate	0.1–0.4	g/s
Clad layer thickness	0.1–2	mm
Feeder nozzle diameter	3	mm
Feeder nozzle orientation	30 with substrate	°
Shielding gas	Argon	—
Overlap between passes	50	%

4. Wire or ribbon feed.
5. Powder feed.

Other methods of coating include thermal or plasma spraying, screen printing, and electroplating. The listed methods will now be further discussed in the following sections.

11.3.2.1 Preplaced Powder With this method (Figure 11.18), the powder, either in loose form or in a slurry mixture, is placed on the substrate, and the laser beam is directed to melt through the powder and fuse it to the substrate. The major drawbacks of this method are listed here.

1. It is limited to flat beds. Generally, complex geometries are difficult to clad by this approach.
2. Shielding gas usage is difficult, as it tends to blow away the powder.

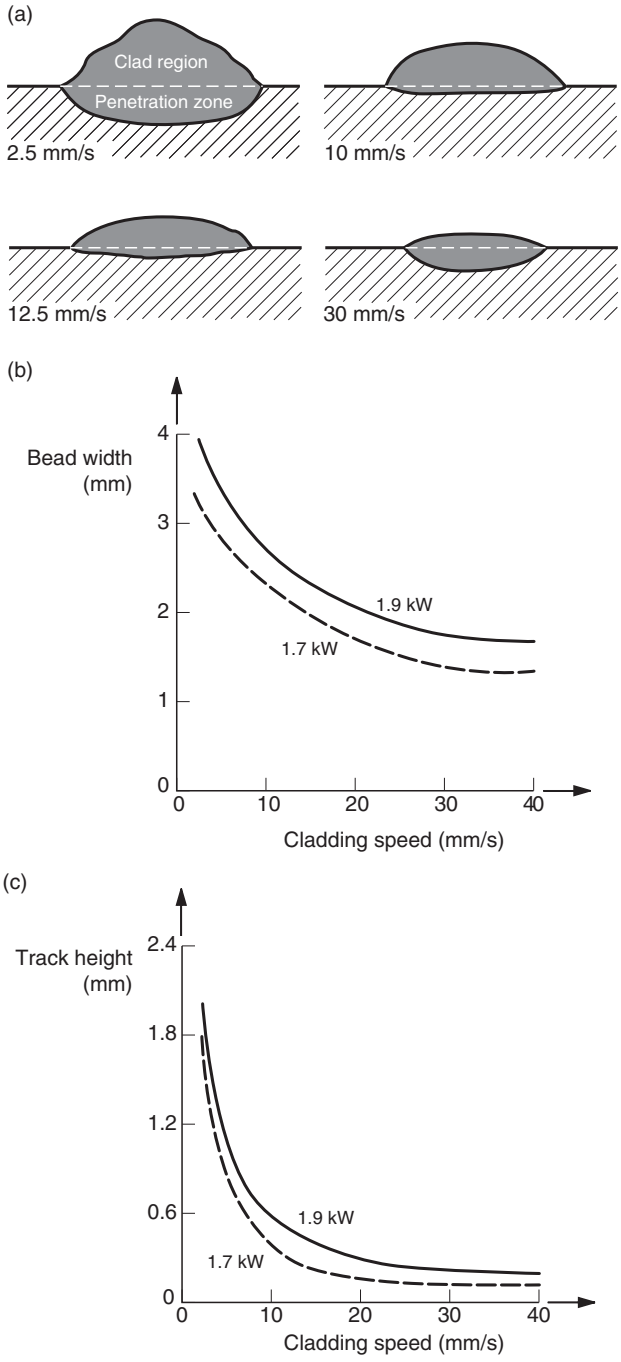


Figure 11.17 Effect of traverse speed on clad configuration. (a) Cross-sectional view of clad track. (b) Variation of bead width with cladding speed. (c) Variation of track height with cladding speed. Source: Jasim and West (1989)/SPIE.

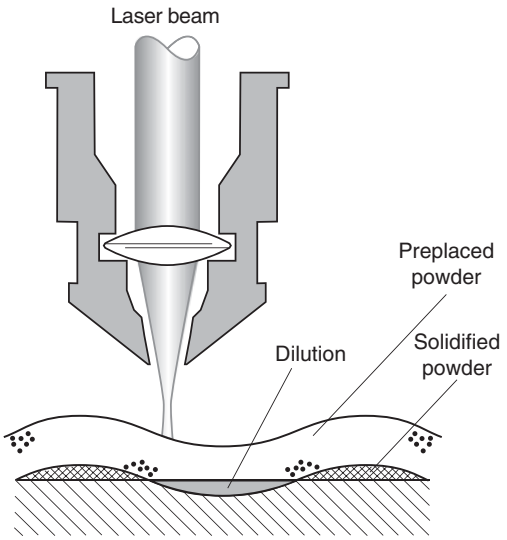


Figure 11.18 Preplaced powder method of laser cladding. Source: Eboo and Lindemanis (1985)/SPIE.

3. Evaporation or decomposition of the binder may lead to porosity.
4. Control of the molten zone between the clad layer and substrate is difficult. Variations in bed depth may either cause excessive dilution (small depth) or absence of metallurgical bonding (greater depth).
5. It requires a higher specific energy to melt through a specified bed depth than that for a powder feed system.

However, it has the advantage that it is very simple to use and is also relatively low cost.

11.3.2.2 Presprayed Surfacing Powder This method involves first spraying surfacing powder on the substrate by thermal or plasma spraying, and subsequently consolidating by laser melting. The additional processing step involved is a drawback of the process.

11.3.2.3 Prepositioned Chip A chip that is made to the shape of the component's surface is placed on the substrate and consolidated by laser melting to form a metallurgical bond. The setback of this approach is also the extra processing step.

11.3.2.4 Wire Feed This process is similar to arc or gas welding. In this case, the wire is fed into a laser generated melt zone at the substrate. It involves a single step, and complex shapes can be clad.

11.3.2.5 Powder Feeding The setup for this method consists of the laser, powder delivery system, and shielding gas (Figure 11.16). It involves continuously feeding the alloy powder through a nozzle that is positioned such that the powder feeds directly into the molten portion of the substrate. Different powder compositions can be fed in, enabling online variations in the alloy composition to be achieved.

The powder may be dispensed in one of two ways:

1. Gravity feeding, where the powder is delivered directly in front of the molten zone, and in that respect, resembles a preplaced powder bed. It may lead to segregation of widely different density particles.

2. Using a pneumatic screw feed system with a small amount of argon gas (about 1 l/min) flowing in the tube to aid powder flow and directional placement. Control of the powder flow rate into the molten zone is determined by the following factors:
   (a) Material flow characteristics.
   (b) Powder preparation, for example, preheat.
   (c) Powder injection angle.
   (d) Shielding gas characteristics.

To reduce oxidation and/or elemental loss, the process may be carried out in a slightly pressurized inert enclosure, with the laser beam being introduced through a NaCl window.

Since the cladding process involves addition of one or more materials to the substrate, this results in dilution or change in chemical composition of portions of the base metal.

11.3.3 Dilution

Dilution of the clad material can be defined either as the ratio of the area of substrate melted to the total melted zone or in terms of the chemical composition difference between the starting powder and the solidified clad layer (Figure 11.19). These may be expressed as

$$\%\text{Area dilution} = \frac{\text{Melted substrate cross-sectional area}}{\text{Clad region cross-sectional area} + \text{Melted substrate cross-sectional area}} = \frac{(2)}{(1)+(2)} \tag{11.15}$$

which is strictly applicable only to single track clads, and

$$\%\text{Chemical dilution} = \frac{C_{\text{clad}} - C_{\text{powder}}}{C_{\text{powder}}} \tag{11.16}$$

where C_{clad} is the elemental composition of the clad at a position near the substrate/clad interface and C_{powder} is the elemental composition of the starting powder.

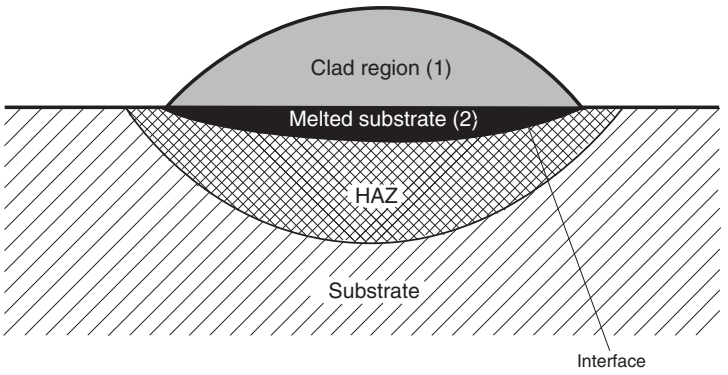


Figure 11.19 Dilution during the cladding process. Source: Jasim and West (1989)/SPIE.

With increasing traverse speed, the area dilution tends to decrease to a minimum and then begin to increase, while the chemical dilution remains constant. The chemical dilution is a measure of the stoichiometry of the alloy formed by cladding, and thus is not expected to change significantly with processing conditions once melting occurs. However, the area dilution is a measure of the amount of substrate that melts, and thus is significanly influenced by the processing speed. A low value of area dilution may be appropriate for say, a hard coating application where it is desired to have a hard layer on the surface of the base material. On the other hand, where a different alloy composition is desired, then a high value of the chemical composition would be desirable.

11.3.4 Advantages and Disadvantages of Laser Deposition

11.3.4.1 Advantages Laser deposition has the following advantages over other more traditional methods:

1. Relatively low heat input, and thus narrow heat affected zone and reduced distortion, with minimal machining being required.
2. Reduced alloy material loss.
3. Minimal clad dilution of base metal (less than 2%) enables the special properties of the clad material to be maintained.
4. Flexibility and ease of automation.
5. Complete metallurgical bonding to the substrate, resulting in high integrity coating.

11.3.4.2 Disadvantages The primary disadvantages of laser deposition are as follows:

1. It is not the method of choice when the area to be hardened is relatively small.
2. It is more expensive than comparative processes that use plasma arc or gas tungsten arc heat sources.

In general, laser deposition is an effective means of changing the microstructure and thus mechanical properties of thin layers of a substrate where the surface area to be processed is large enough. Other methods of treatment exist that do not involve melting of the substrate, for example, laser physical vapor deposition and laser chemical vapor deposition. There are a number of other surface modification techniques that do not involve the use of lasers. However, in the following Section 11.1, we further discuss the laser physical vapor deposition technique.

11.4 LASER PHYSICAL VAPOR DEPOSITION (LPVD)

The principle of LPVD is illustrated in Figure 11.20. A CW laser beam is irradiated on a target in a vacuum chamber. The target is placed on a traveling holder and inclined to the beam axis, while the substrate is fixed and held parallel to the target. The target is moved at a preset speed, and the substrate may be preheated. The laser beam vaporizes material from the target, which is deposited on the substrate. As an illustration, typical key processing parameters would be as shown in Table 11.4.

The deposition rate increases with an increase in laser power, but decreases as the target–substrate distance increases (Figure 11.21a). On the other hand, the resulting hardness decreases with an

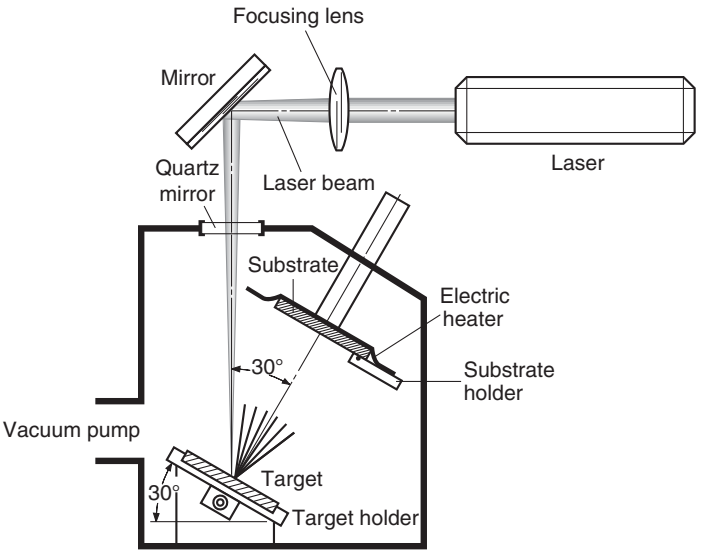


Figure 11.20 Schematic of the laser physical vapor deposition process. Source: From *ICALEO 1991 Proceedings*, Vol. 74, Copyright 1992, Laser Institute of America. All rights reserved.

Table 11.4 Typical process parameters used in laser physical vapor deposition.

Parameters	Values	Units
Power	100–200	W
Scanning velocities	3	mm/s
Chamber pressure	1×10^{-4}	Pa
Target orientation	30	°
Target–substrate distance	50	mm
Substrate temperature	800	K

increase in laser power and increases as the target–substrate distance increases (Figure 11.21b). The hardness also increases with an increase in substrate temperature.

11.5 LASER SHOCK PEENING

Laser shock peening is a process that imparts compressive residual stresses on the surface of a material. It is used primarily for increasing fatigue life, corrosion resistance, and improving resistance to cracking since the compressive stresses retard the growth of cracks initiated on the surface.

It is normally done using a thermoprotective coating or absorbing layer of black paint or tape applied to the surface of the part (Figure 11.22). The coating also serves to protect the substrate from the thermal effects of the ablation and further enhances the magnitude of the resulting shock wave pressure by forming a plasma. A layer of dielectric material which is transparent to the laser beam, usually glass or water, is placed on top of the absorbing layer. The laser beam instantly vaporizes the absorbing layer and produces plasma that expands rapidly, creating very high pressures

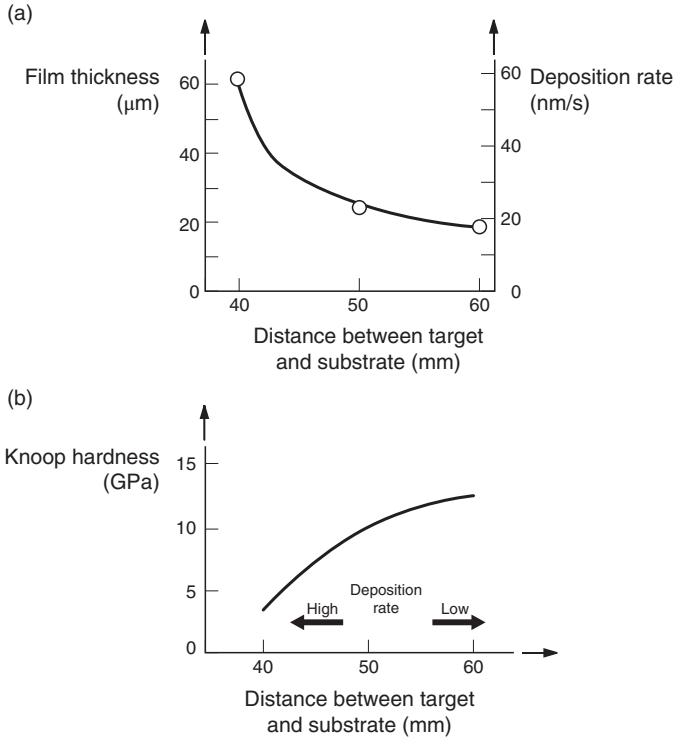


Figure 11.21 Effect of target–substrate distance on film thickness and hardness for a nickel substrate at a laser power of 100 W, substrate temperature of 673 K, and a pressure of 2.6×10^{-3} Pa. (a) Film thickness. (b) Film hardness. Source: From *ICALEO 1991 Proceedings*, Vol. 74, Copyright 1992, Laser Institute of America. All rights reserved.

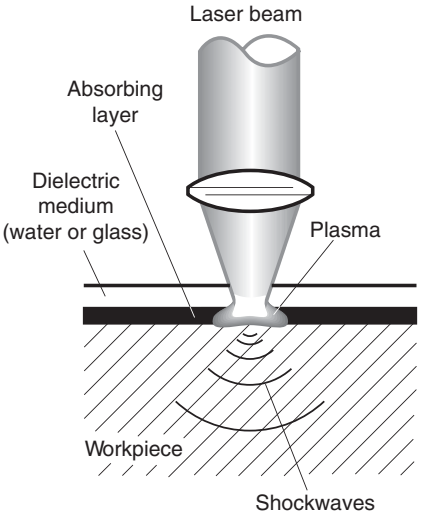


Figure 11.22 Schematic of the laser shock peening process.

at the interface, as a result of the recoil momentum of the ablated material. The pressure results in shock waves which are transmitted throughout the material, and induces compressive stresses in the material. Plastic deformation occurs when the compressive stress exceeds the Hugoniot elastic limit (HEL). With the plasma confined by the layer of dielectric material, pressures of the order of gigapascals are obtained. Pressures generated in the confined mode are typically an order of magnitude greater than in direct ablation, for similar laser irradiances. Also, using quartz as the overlay material results in much higher pressures than when water is used, since the acoustic impedance of water is about 10% that of quartz. However, water is more flexible to use and thus suited for complex geometries. Furthermore, its threshold for dielectric breakdown is higher than that of quartz.

Use of the thermoprotective coating avoids melting of the metallic surface during the irradiation. Without the coating, the melting that results would create tensile residual stresses. The process requires a laser system with high average power and large pulse energy for a high throughput. This is normally achieved with a Q-switched laser operating in the near infrared wavelength range which is effectively transmitted by the water or glass overlay material, for example a Nd:Glass or Nd:YAG laser with a wavelength of 1.06 μm. Even though using a shorter wavelength laser, for example by frequency doubling to 0.53 μm, would increase the shock wave generation due to enhanced photon–metal interaction, this is not recommended for a number of reasons. The process of frequency doubling can reduce the pulse energy by as much as 50%. Furthermore, at the shorter wavelength, the threshold for dielectric breakdown of the surrounding atmosphere is reduced, resulting in more plasma being generated above the workpiece surface, and that absorbs part of the incoming radiation (inverse bremsstrahlung absorption). Typical process parameters are listed in Table 11.5.

Laser beams with a circular profile are often used for this process, but have the disadvantage that the resulting stress distribution is not uniform. Circular profiles produce a stress dip in the center, which can be a preferred location for crack initiation. For these reasons, a square profile is preferable. The percent of cold work and compressive residual stress increase as the number of pulses increases, with the magnitude of the compressive stress being proportional to the number of pulses per spot. The disadvantages of laser peening are that it is a relatively slow process; it can also be quite expensive, as layers of absorbing material must be applied before each peening application. However, it is becoming a more viable alternative to shot peening as new developments in laser systems allow faster processing.

Laser shock peening has traditionally been done using nanosecond pulse lasers. However, recent work has demonstrated the potential for using femtosecond pulse lasers. As an illustration, a 2024-T351 aluminum alloy that was shock peened using a 120 fs laser of 800 nm wavelength without any absorbing or confining layer (dry laser peening) resulted in a much lower surface roughness than that performed using a nanosecond laser. Furthermore, compressive residual stresses were measured for depths up to 100 μm, along with hardness values that were about twice those of

Table 11.5 Typical process parameters used in laser shock peening.

Parameter	Value	Units
Power (average)	1	kW
Pulse power density	1–50	GW/cm^2
Pulse energy	100	J
Pulse duration	10^{-4}–50	ns
Pressure generated	1–20	GPa
Residual stress depth	0.5–2	mm

the unpeened aluminum alloy. In addition, dry laser peening resulted in significant improvements in fatigue life compared to that of the base material. One advantage with dry laser peening is that it eliminates the time and cost associated with coating application and removal.

11.5.1 Background Analysis

The pressure generated at the interface can be analyzed using the simplified set-up shown in Figure 11.23, considering glass to be the dielectric material. The pressure results in shock waves in both the workpiece and the glass overlay and causes the interface walls to be displaced with velocities u_1 and u_2.

If the thickness of the interface at time t is $L_p(t)$, then we have

$$L_p(t) = \int_0^t [u_1(t) + u_2(t)]dt \tag{11.17}$$

The velocities, u_i ($i = 1, 2$), are related to the pressure P by

$$P = \rho_i U_i u_i = Z_i u_i \tag{11.18}$$

where $i = 1, 2$ is the material 1 or 2, U_i is the shock velocity (m/s), Z_i is the shock impedance (g/cm^2/s), and ρ_i is the density (kg/m^3).

For the situation where the two materials are solids, we can assume that their shockwave impedances, $Z_i = \rho_i U_i$, are constant. Then from Eqs. (11.17) and (11.18), we get

$$\frac{dL_p(t)}{dt} = \frac{2}{Z}P(t) \tag{11.19}$$

where $P(t)$ is the plasma pressure at time t, and $\frac{2}{Z}$ is $\frac{1}{Z_1} + \frac{1}{Z_2}$.

The laser energy that is absorbed increases the internal energy per unit volume, Q_e, of the interface plasma and also does work to open up the interface. If the laser intensity is $I(t)$, then the energy per unit surface area deposited over the time interval dt is $I(t)\,dt$. Let the plasma thickness increase by dL_p during this time interval. The increase in internal energy is then given by $d[Q_e(t)L_p]$ (Joules per unit area), and the work done per unit area by the forces resulting from the pressure is $P(t)\cdot dL_p$. Thus,

$$I(t) = P(t)\frac{dL_p(t)}{dt} + \frac{d[Q_e(t)L_p(t)]}{dt} \tag{11.20}$$

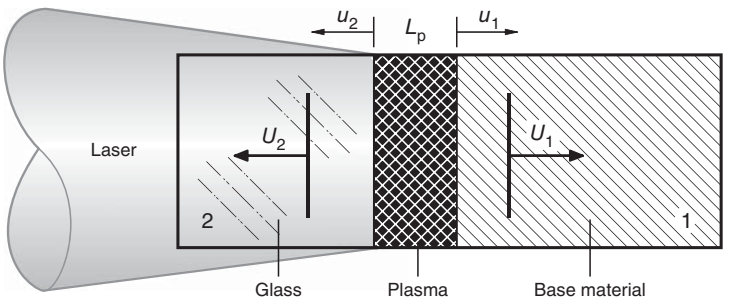


Figure 11.23 Schematic of the confined ablation process. Source: Adapted from Fabbro et al. (1990).

We now make the following assumptions:

1. The fraction, f_E, of the internal energy, $Q_e(t)$, which represents the thermal energy, $Q_T(t)$, is constant. The balance, $1-f_E$, is used to ionize the gas. f_E is typically ≈0.1–0.2.
2. The interface plasma obeys the ideal gas laws.

The relationship between the thermal energy, Q_T, of an ideal gas, and the resulting pressure, P, is given by

$$P(t) = \frac{2}{3}Q_T(t) = \frac{2}{3}f_E Q_e(t) \tag{11.21}$$

Equation (11.20) can then be rewritten as

$$I(t) = P(t)\frac{dL_p(t)}{dt} + \frac{3}{2f_E}\frac{d[P(t)L_p(t)]}{dt} \tag{11.22}$$

Further assume that the laser beam intensity, $I(t) = I_0 =$ constant, and that $L_p(0) = 0$. Then the pulse duration is $\Delta\tau_p$, Eqs. (11.19)–(11.22) can be solved to give the plasma pressure at the interface, P (which is then constant), as

$$P = 0.10 \times \sqrt{\left[\frac{f_E}{2f_E + 3}\right] ZI_0} \tag{11.23}$$

and the resulting plasma thickness is

$$L_p(\Delta\tau_p) = 2 \times 10^4 \frac{P\Delta\tau_p}{Z} \tag{11.24}$$

where the parameters have the following units:

P in kbars, Z in g/cm^2 s, I_0 in GW/cm^2, L_p in μm, and $\Delta\tau_p$ in ns.

If, on the contrary, $L_p(0) = L_{p0} \neq 0$, then $P(t)$ is not constant. Under these circumstances, it can be shown, using Eq. (11.19) and neglecting the term $P(t)\frac{dL_p(t)}{dt}$ in Eq. (11.22), that

$$L_p(t)^2 = {L_{p0}}^2 + \left(\frac{2P_0 t}{Z}\right)^2$$

The pressure, which is then time dependent, becomes

$$P(t) = P_0\left(1 - \frac{{L_{p0}}^2}{L_p(t)^2}\right)^{1/2} \tag{11.25}$$

where $P_0 = 0.10 \times \sqrt{\left[\frac{f_E}{3}\right] ZI_0}$.

Under the high impact loading associated with laser shock peening, the material becomes plastically deformed when the pressure (or compressive stress due to the pressure) reaches the HEL, which may be expressed as follows:

$$\text{HEL} = \left(\frac{1-\nu_p}{1-2\nu_p}\right)\sigma_{yd} \tag{11.26}$$

where σ_{yd} is the dynamic yield strength of the material at high strain rates, and ν_p is the Poisson ratio.

For shock wave pressures greater than twice the HEL, the maximum residual stress, σ_{rs}, induced on the surface may be estimated by

$$\sigma_{rs} = -\frac{P_{sw}}{2\left(1+\frac{\lambda_L}{2\mu_L}\right)}\left[1-\frac{4\sqrt{2}}{\pi r}(1+\nu_p)\frac{u_{el}u_{pl}\Delta\tau_p}{u_{el}-u_{pl}}\right] \tag{11.27}$$

And the depth of the residual stress distribution, d_{rs}, is given by

$$d_{rs} = \left[\frac{u_{el}u_{pl}\Delta\tau_p}{u_{el}-u_{pl}}\right] \tag{11.28}$$

where P_{sw} is the pressure due to the shock wave; λ_L and μ_L are the elastic Lame constants of the workpiece; r is the radius of the impact; u_{el} and u_{pl} are the respective elastic and plastic velocities of the shock wave; and $\Delta\tau_p$ is the duration of the pressure pulse.

Increasing the number of laser pulses at a given location increases the depth of induced stresses.

Example 11.4 In shock peening the surface of an aluminum plate, the laser that is used generates a pulse energy of 80 J at a pulse duration of 10 ns, and a beam diameter of 3 mm. If the plasma thickness obtained at the end of the pulse is 40 μm, determine the interface pressure generated. Assume that the pressure is invariant with time, that 15% of the internal energy is expended as thermal energy, and that the pulse energy is uniform throughout the pulse duration.

Solution:
The plasma thickness is given by Eq. (11.24):

$$L_p(\Delta\tau_p) = 2\times10^4\frac{P\Delta\tau_p}{Z}$$

$$\Rightarrow Z = 2\times10^4\frac{P\Delta\tau_p}{L_p(\Delta\tau_p)} = 2\times10^4\times\frac{P\times10}{40} = 5\times10^3\times P$$

Thus, from Eq. (11.23), we have

$$P = 0.10\times\sqrt{\left[\frac{f_E}{2f_E+3}\right]ZI_0} = 0.10\times\sqrt{\left[\frac{f_E}{2f_E+3}\right]\times5\times10^3\times P\times I_0}$$

But the pulse power, q, is given by

$$q = \frac{80}{10\times10^{-9}} = 8\times10^9\ \text{W}$$

and

$$I_0 = \frac{8\times10^9}{\pi\times0.3^2/4} = 113\ \text{GW/cm}^2$$

Thus,

$$P = 0.10\times\sqrt{\left[\frac{0.15}{2\times0.15+3}\right]\times5\times10^3\times P\times113}$$

$$\Rightarrow P = 256.8\ \text{ kbars} = 25.7\ \text{ GPa}$$

11.5.2 Thermal Relaxation at High Temperatures

Thermal relaxation is discussed separately since laser shock peening is often applied to products that are used in high temperature environments such as turbine blades. At such temperatures, there is a greater relaxation of the induced stresses and that may result in loss of some or all the benefits associated with the peening process. However, stress relaxation tends to produce a more uniform spatial stress distribution.

The cold working associated with shot peening tends to be quite severe, but yet confined to shallow depths of about 75–250 μm. The severe deformation makes the material unstable, resulting in rapid loss of the compressive residual stresses at relatively high temperatures.

The major mechanisms responsible for residual stress relaxation at high temperatures include the following:

a. Creep-controlled rearrangement of dislocation network.
b. Annihilation of meta-stable defects.
c. Softening of the material.

The extent of relaxation increases with the amount of initial cold work induced, exposure temperature, and time of exposure at the high temperature. Most of the relaxation occurs in the initial stages, eventually leveling off. It is also highest at the surface where the cold work is most severe and reduces with depth into the workpiece.

The relaxation process may be analyzed using a relationship that is based on the Zener–Wert–Avrami model:

$$\frac{\sigma_{\mathrm{rs}}}{\sigma_0} = e^{-(Dt)^n} \tag{11.29}$$

where σ_{rs} is the residual stress at time t, σ_0 is the initial stress before the start of heating, n is a constant that depends on the dominant relaxation mechanism, and D is a coefficient that has an Arrhenius temperature behavior:

$$D = D_0 e^{-\left(\frac{Q_{\mathrm{d}}}{k_{\mathrm{B}}T}\right)} \tag{11.30}$$

such that D_0 is a constant, Q_{d} is the activation energy for relaxation, k_{B} is the Boltzmann constant, and T is the absolute temperature.

On the contrary, the cold working induced by laser shock peening is relatively small compared to that associated with shot peening, but penetrates much deeper below the surface (about 1–2 mm). Thus, the residual stress relaxation that results after exposure to heat is much smaller for parts that had been subjected to laser shock peening. For example, in the titanium alloy Ti_6Al_4V, very little relaxation occurs and in Inconel 718, almost no relaxation is observed at temperatures up to 700 °C. However, 50% relaxation occurs in 10 minutes for the same materials if they are subjected to shot peening before heat treatment at 700 °C.

11.5.3 Advantages and Disadvantages of Laser Shock Peening

11.5.3.1 Advantages Laser shock peening has several advantages over traditional shot peening, and these are listed as follows:

1. It produces compressive residual stresses that extend deeper into the material (over 2 mm) than can be obtained with shot peening.
2. The compressive residual stresses are also produced with less cold work than with shot peening, allowing for less thermal relaxation of these stresses when the part is subjected to high temperatures.

3. With conventional shot peening, the surface of the workpiece is dimpled, which can be undesirable esthetically. Laser shock peening does not adversely affect the surface appearance or quality.
4. Since the process is noncontact, it is suitable for complex geometries and areas that are difficult to access, for example, notches and small fillets.

11.5.3.2 Disadvantages The major disadvantage of laser shock peening is that it currently requires specially designed lasers for production runs, since it requires a laser system with high average power and large pulse energy for a high throughput.

11.5.4 Applications

Laser shock peening has applications in the aerospace, automotive, nuclear, orthopedic, and power industries. One application of the process has been in making the leading edge of fan blades resistant to foreign object damage. Others include turbine blades for thermal power plants, bearing components, gear shafts in the aerospace industry, and automotive components such as leaf springs.

11.6 LASER TEXTURING

Surface texturing is a means of modifying the geometric features of a surface to impart desired properties and functionality to it. The properties may involve minimized surface reflectivity (light trapping, which may, for example, result in the formation of black silicon by enhancing the absorptivity of silicon in the visible spectrum), enhanced tribological characteristics, modified fluid flow characteristics, and super wetting surfaces. There are different methods for texturing a surface. It can be done by mechanical means, by chemical etching, with reactive ion etching, using photolithography, or with lasers. Our focus in this section is the use of lasers for surface texturing.

Lasers have the advantage of being noncontact, making it easier to impart different geometric features unto a surface. These features can be at the microscale or at the nanoscale. Creation of nanofeatures is discussed under nanotechnology applications in Section 14.2. Here we focus on applications that involve micromachining of specific geometries unto the surface. Two common approaches involve scribing grooves or creating micropores to generate different patterns. In either case, the quality of the surface generated is influenced by the laser processing parameters. Of particular importance is the pulse duration.

Pulse duration in the nanosecond range or more (typical of Nd:YAG lasers) is longer than the electron–phonon thermal relaxation time (of the order of picoseconds). This is the time it takes to transfer the energy absorbed by electrons to the lattice. Thus, the electrons and lattice are essentially in thermal equilibrium at the end of a long pulse, and material removal then involves both melting and vaporization, resulting in the formation of recast layers that affect the surface texture. The recast layer (ripple) on the periphery of a micropore or along the edges of a groove is the result of outward flow of molten metal (Figure 9.26). This can have significant influence on the surface texture generated and thus the performance of the surface.

Figure 11.24 illustrates a typical micropore formed with an Nd:YAG laser of pulse duration in the range 100 to 600 ns and peak power of 80 W on a flat $100Cr_6$ steel plate to improve tribological performance of contact surfaces. The geometry of the micropore (depth, diameter, and ripple height) is determined by the processing parameters such as beam mode, energy per pulse, and number of pulses.

A pattern of grooves generated on a multicrystalline silicon wafer using an Nd:YAG laser to enhance the conversion efficiency of solar cells is shown in Figure 11.25 and further demonstrates the formation of a recast layer along the groove edges. This reduces the efficiency of the cells. Further etching to remove the layer improves the efficiency, but then increases the cell reflectance.

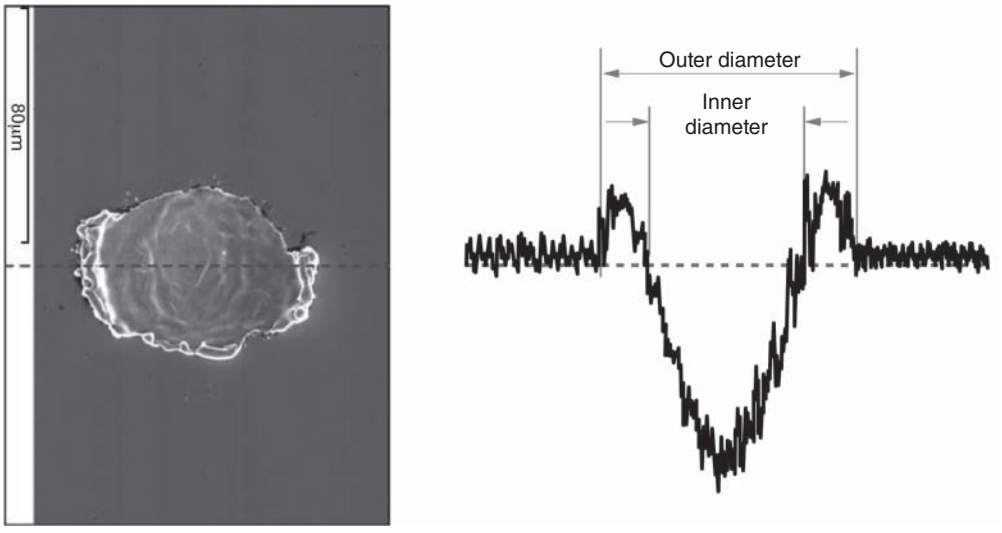


Figure 11.24 (a) SEM micrograph and (b) 2D profile illustrating the geometry of a micro-pore. Source: Vilhena et al. (2009)/Elsevier.

As previously discussed in Section 9.3.1 under micromachining, ultrashort pulse beams have a shorter duration than the electron–phonon thermal relaxation time. The material is thus directly vaporized, with very little or no melting, resulting in cleaner micropores or grooves (Figure 9.26). One such application of femtosecond lasers in surface texturing involves generating a surface geometry that enables conversion of boundary layer flow in the heated regions of aircraft and rocket engines from laminar to turbulent flow to enhance cooling of the components. Using a 0.775 μm wavelength femtosecond laser with 180 fs pulse duration and a focused beam size of 30 μm on the ultrafine polished surface of nickel-based superalloys, ripples with periodicity of the order of the laser wavelength are obtained on the surface at low average power densities of about 0.33 W/cm^2. Such ripples are generally referred to as laser-induced periodic surface structures (LIPSS) and result from interference

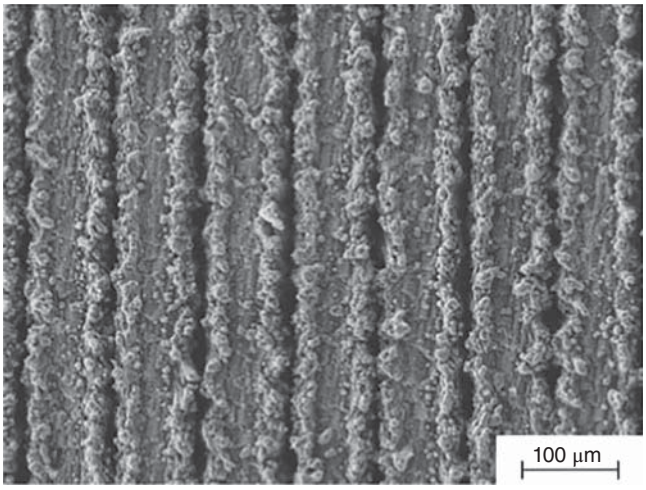


Figure 11.25 SEM micrograph of grooves generated on a multi-crystalline silicon wafer using an Nd:YAG laser. Source: Dobrzański and Drygała (2008)/International OCSCO World Press.

between the incident beam and surface polaritons, which generate a standing wave on the surface that results in selective material ablation (see also Section 14.2.1). LIPSS can be either low spatial frequency (LSFL) or high spatial frequency (HSFL). LSFLs are common with most materials, while HSFLs tend to be associated with semiconductors and dielectrics, and have spatial periods that are much smaller than the laser wavelength. LIPSS has been shown to result in friction reduction during sliding.

At higher power densities (higher fluence and lower scanning speed), micropores that are aligned in directions perpendicular to the ripples begin to form, resulting in higher surface roughness. To achieve a given ablation depth at lower power densities (lower fluence and higher scanning speed), more scans are needed than would be needed at higher power densities. However, a smaller surface roughness is associated with the lower power density.

11.7 SUMMARY

In this chapter, we have learnt that laser surface modification is used to change either the surface composition or microstructure of a material to give it certain desired properties, and the principal techniques include the following:

Laser Surface Heat Treatment: This involves exposing the surface of a material to a thermal cycle of rapid heating and cooling such that the surface layers are hardened, either through martensitic transformation in the case of steels or through precipitation hardening, in the case of nonferrous alloys. The process does not involve melting, and the transformations occur in the solid state.

Laser Surface Melting (*Skin Melting, or Glazing*): In this technique, a thin surface layer of material is melted with the laser beam and subsequently undergoes rapid solidification as a result of self-quenching, resulting in alterations in the local microstructure. The process is similar to autogenous (no filler metal added) conduction mode laser welding, but the resulting cooling rates are high enough to induce rapid solidification. This process is especially suited to materials that are generally difficult to harden by solid-state transformation, for example, steels with widely dispersed carbide or graphite (mainly low-carbon steels). The microstructural changes may be accompanied by changes in properties such as hardness, corrosion resistance, and wear resistance.

Laser Deposition (*Cladding, Alloying, and Hard Coating*): This involves applying a second material to the surface and melting it with a laser beam either to alloy with a thin surface layer of the base material or to bond to the surface. As in the other cases of laser surface modification, the rapid solidification and/or quench rates associated with the process enable either amorphous or nonequilibrium crystalline phases to be formed. The major methods for applying the coating include preplaced loose powder bed; presprayed surfacing powder; prepositioned chip or rod; wire or ribbon feed and powder feed.

Laser Physical Vapor Deposition: In this method, the second material is applied in vapor form, and no melting takes place. A CW laser beam is irradiated on a target in a vacuum chamber and vaporizes material from the target, which is deposited on the substrate, forming a thin layer. The deposition rate increases with an increase in laser power, but decreases as the target–substrate distance increases. The resulting hardness, on the other hand, decreases with an increase in laser power and increases as the target–substrate distance increases.

Laser Shock Peening: This involves using the laser beam to vaporize a thin layer of thermoprotective coating or absorbing layer of black paint or tape. This is placed beneath a layer of dielectric material that is transparent to the laser beam, usually glass or water. The recoil momentum of the ablated material results in high pressure that generates shock waves, inducing compressive residual stresses on the surface of the material. It is used primarily for

increasing fatigue life and improving resistance to cracking. One issue with shock peened parts is that stress relaxation occurs when subjected to relatively high temperatures.

Laser Surface Texturing: With this process, the geometry of the workpiece surface is modified using a laser beam, to induce some desirable properties on the surface. Such properties may include enhanced tribological, surface reflectivity, and fluid flow characteristics.

APPENDIX 11.A

Symbol	Parameter	Units
a	Half length of resonator mirror	m
A_1	Eutectoid temperature	°C
A_3	Lower temperature boundary of austenite zone	°C
c_d	Constant	—
C_a	Carbon composition in % by weight	%
C_c	Critical carbon fraction above which martensitic transformation occurs	%
C_{clad}	Elemental composition of the clad at a position near the substrate/clad interface	%
C_e	Carbon content of pearlite	%
C_f	Carbon content of ferrite	%
C_{powder}	Elemental composition of the starting powder	%
d_{rs}	Depth of residual stress distribution	mm
D	Stress relaxation coefficient	/s
D_a	Diffusion coefficient for carbon	mm^2/s
D_0	Diffusion constant	mm^2/s
f	Volume fraction of martensite formed at time t	—
f_E	Fraction of internal energy which represents the thermal energy	—
f_i	Volume fraction initially occupied by pearlite colonies	—
f_m	Maximum volume fraction permitted by phase diagram	—
g_s	Average austenite grain size	μm
HEL	Hugoniot elastic limit	MPa
l_p	Spacing of pearlite plates in a colony	μm
$L_p(t)$	Thickness of plasma interface at time t	μm
L_{p0}	Initial thickness of plasma interface at time $t = 0$	μm
n	Constant	—
P_0	Initial plasma pressure	kbar (100 MPa)
P_{sw}	Pressure due to shock wave	MPa
Q_a	Activation energy for transformation	J
Q_d	Activation energy for relaxation	J
Q_e	Internal energy of the interface plasma	J
Q_T	Thermal energy of the interface plasma	J
r	Radius	mm
t_0	Time to diffuse a distance equal to beam radius	s

Symbol	Parameter	Units
u_{el}, u_{pl}	Elastic and plastic velocities of the shock wave	m/s
u_i	Velocity of medium i	m/s
u_x	Traverse velocity of the heat source in the x-direction	m/s
U_i	shock velocity of medium i	m/s
z_0	Distance over which heat can diffuse during the beam interaction time $\frac{w}{u_x}$	mm
$Z = 2\frac{Z_1 Z_2}{Z_1+Z_2}$	Overall shock impedance	g/cm^2 s
Z_i	Shock impedance of medium i	g/cm^2 s
λ_L and μ_L	Elastic Lame constants	MPa
ρ_i	Density of medium i	kg/mm^3
σ_0	Initial stress	MPa
σ_{rs}	Residual stress	MPa
σ_{yd}	Dynamic yield stress	MPa
τ_h	Thermal time constant	s
$\Delta\tau_p$	Pulse duration	ns
θ_c	Semi angle subtended by mirrors at each other's center	°
θ_d	Semi angle of diffracted waves	°

APPENDIX 11.B

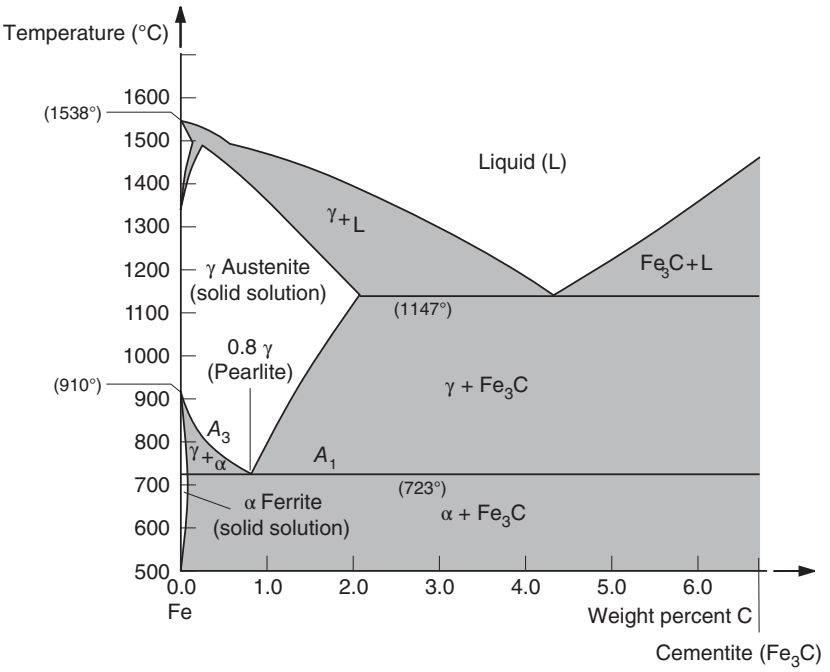


Figure B11.1 The iron–carbon diagram.

PROBLEMS

11.1. The surface of an AISI 1080 steel gear is to be heat treated using a 2.5 kW Nd:YAG laser. The purpose is to enhance the wear resistance. The defocused beam diameter is 5 mm, and the scan rate is set at 10 mm/s. Only a single pass of the laser beam will be used for the heat treatment. Determine the time, t, required for the diffusion controlled transformation of pearlite to austenite (pearlite dissolution) that results during the process.

Assume:

(1) The temperature is time-dependent
(2) The diffusion coefficient for carbon, $D_a = 9 \times 10^{-10}$ m^2/s
(3) The diffusion constant, D_0, for the diffusion of carbon in α-Fe (ferrite) is 0.020 cm^2/s = 2×10^{-6} m^2/s (Shewmon, Table 2-1, p. 64).
(4) The activation energy for diffusion, Q_a, is 20.1 kcal/mol = 84.15 kJ/mol
(5) The reflection coefficient of the surface is 80%
(6) Ambient temperature is 25 °C.

11.2. For the conditions in Problem 11.1, determine the average size of the zone over which martensite can be formed.

11.3. For the conditions in Problem 11.1, determine the volume fraction of martensite that can be formed. You may assume that the average austenite grain size is 200 μm.

11.4. A 7 kW CO_2 laser is used to heat treat the surface of an AISI 1045 steel cam at a beam diameter of 6 mm, and a scan rate of 25 mm/s. Determine the spacing of the pearlite plates that will enable the pearlite colonies to be fully transformed to austenite under these conditions.

Assume:

(1) The temperature is time-dependent
(2) The diffusion constant, D_0, for the diffusion of carbon in α-Fe (ferrite) is 0.020 cm^2/s = 2×10^{-6} m^2/s (Shewmon, Table 2-1, p. 64)
(3) The activation energy for diffusion, Q_a, is 20.1 kcal/mol = 84.15 kJ/mol (Shewmon, Table 2-1, p. 64)
(4) The reflection coefficient of the surface is 85%
(5) Ambient temperature is 25 °C.

11.5. The leading edges of titanium turbine blades are to be shock peened using a laser that generates a pulse energy of 100 J with a pulse duration of 15 ns, and a beam diameter of 4 mm. If a plasma of thickness 50 μm develops by the time the pulse is over, determine the pressure that is generated at the interface between the substrate and the dielectric material. Assume that the pressure is invariant with time; that 10% of the internal energy is expended as thermal energy; and that the pulse energy is uniform throughout the pulse duration.

11.6. The interface impedance associated with a shock peening process is determined to be 0.6×10^6 g/cm^2 s. If the laser being used is capable of generating a pulse energy of 95 J, determine the pulse duration necessary for generating a plasma thickness of 60 μm at the end of the pulse. What is the resulting interface pressure? Assume that the beam diameter is 5 mm; that the pressure is invariant with time; that 12.5% of the internal energy is expended as thermal energy; and that the pulse energy is uniform throughout the pulse duration.

11.7. Explain how preheating affects the absorption of a material during laser surface heat treatment.

11.8. An AISI 1060 steel is subjected to laser heat treatment. The working conditions are as follows: Nd:YAG laser output power = 4.5 kW; defocused beam diameter = 5 mm; scan rate (u_x) = 8 mm/s; reflection coefficient of the surface = 90%; ambient temperature, $T_0 = 25\,°C$; diffusion constant, $D_0 = 2\,mm^2/s$; average austenite grain size, $g_s = 200\,\mu m$; activation energy for diffusion, $Q_a = 20.1$ kcal/mol.

Calculate:

(a) The spacing of the pearlite plates that will enable the pearlite colonies to be fully transformed into austenite.
(b) The average size of the region where martensite can be formed.
(c) The volume fraction of martensite that can be formed.

BIBLIOGRAPHY

Abbas, G. and West, D.R.F. (1989). Laser produced composite metal cladding. *SPIE, High Power Lasers and Laser Machining Technology* 1132: 232–236.

Ashby, M.F. and Easterling, K.E. (1984). The transformation hardening of steel surfaces by laser beams – I. Hypo-eutectoid steels. *Acta Metallurgica* 32 (11): 1935–1948.

Azuma, H., Takeuchi, A., Ito, T. et al. (2002). Pulsed KrF excimer laser annealg of silicon solar cell. *Solar Energy Materials & Solar Cells* 74: 289–294.

Baeri, P. and Rimini, E. (1996). Laser annealing of silicon. *Materials Chemistry and Physics* 46: 169–177.

Bataille, F., Kechemair, D., and Houdjal, R. (1991). Real time actuating of laser power and scanning velocity for thermal regulation during laser hardening. *SPIE, Industrial and Scientific Uses of High-Power Lasers* 1502: 135–139.

Bradley, J.R. (1986). Experimental determination of the coupling coefficient in laser surface hardening of steel. *SPIE, Laser Processing: Fundamentals, Applications, and Systems Engineering* 668: 23–30.

Bradley, J.R. (1988). A simplified correlation between laser processing parameters and hardened depth in steels. *Journal of Physics D: Applied Physics* 21: 834–837.

Braisted, W. and Brockman, R. (1999). Finite element simulation of laser shock peening. *International Journal of Fatigue* 21: 719–724.

Chahardehi, A., Brennan, F.P., and Steuwer, A. (2010). The effect of residual stresses arising from laser shock peening on fatigue crack growth. *Engineering Fracture Mechanics* 77: 2033–2039.

Dane, C., Hackel, L., Halpin, J. et al. (2000). High-throughput laser peening of metals using a high-average-power Nd:glass laser system. *Proceedings of SPIE - The International Society for Optical Engineering* 3887: 211–221.

Dassow, R., Kohler, J.R., Grauvogl, M. et al. (1999). Laser crystallized polycrystalline silicon on glass for photovoltaic applications. *Solid State Phenomena* 67–68: 193–198.

Dassow, R., Kohler, J.R., Helen, Y. et al. (2000). Laser crystallization of silicon for high-performance thin-film transistors. *Semiconductor Science and Technology* 15: L31–L34.

Ding, K. and Ye, L. (2003a). Three-dimensional dynamic finite element analysis of multiple laser shock peening processes. *Surface Engineering* 19 (5): 351–358.

Ding, K. and Ye, L. (2003b). FEM simulation of two sided laser shock peening of thin sections of Ti–6Al–4V alloy. *Surface Engineering* 19 (2): 127–133.

Dobrzański, L.A. and Drygała, A. (2008). Surface texturing of multicrystalline silicon solar cells. *Journal of Achievements in Materials and Manufacturing Engineering* 31 (1): 77–82.

Eboo, G.M. and Lindemanis, A.E. (1985). Advances in laser cladding process technology. *SPIE, Applications of High Power Lasers* 527: 86–94.

Fabbro, R., Fournier, J., Ballard, P. et al. (1990). Physical study of laser-produced plasma in confined geometry. *Journal of Applied Physics* 68 (2): 775–784.

Fabbro, R., Peyre, P., Berthe, L., and Scherpereel, X. (1998). Physics and applications of laser-shock processing. *Journal of Laser Applications* 10 (6): 265–279.

Fairand, B.P. and Clauer, A.H. (1979). Laser generation of high-amplitude stress waves in materials. *Journal of Applied Physics* 50 (3): 1497–1502.

Fastow, M., Bamberger, M., Nir, N., and Landkof, M. (1990). Laser surface melting of AISI 4340 steel. *Materials Science and Technology* 6 (9): 900–904.

Festa, R., Nenci, F., Manca, O., and Naso, V. (1993). Thermal design and experimentation analysis of laser and electron beam hardening. *ASME Journal of Engineering for Industry* 115: 309–314.

Gnanamuthu, D.S. and Shankar, V.S. (1985). Laser heat treatement of iron-base alloys. *SPIE, Applications of High Powered Lasers* 572: 56–72.

Godbole, V.P. and Chaudhari, S.M. (1988). Laser annealing of silicon. *Bulletin of Materials Science* 11 (2 and 3): 97–108.

Hill, M., DeWald, A., Demma, A. et al. (2003). Laser peening technology. *Advanced Materials and Processes* 161 (8): 65–67.

Hiraoka, H. and Sendova, M. (1994). Laser-induced sub-half-micrometer periodic structure on polymer surfaces. *Applied Physics Letters* 64: 563–565.

Hiraoka, H. (1995). Pulsed UV laser irradiation of polymer surfaces. *Journal of Photochemistry and Photobiology A: Chemistry* 92: 129–133.

Ivarson, A., Powell, J., Ohlsson, L., and Magnusson, C. (1991). Optimisation of the laser cutting process for thin section stainless steels. *ICALEO* 1991: 211–220.

Jasim, K.M. and West, D.R.F. (1989). Laser cladding of carbon steel with a ceramic/metallic composite. *SPIE, High Power Lasers and Laser Machining Technology* 1132: 237–245.

Jasnowski, K.S., Justes, S.P., Amatrian, J.F.Z., and Ibanez, F.G. (1989). Surface hardening: beam shaping and coating techniques. *SPIE, High Power Lasers and Laser Machining Technology* 1132: 246–256.

Jonnalagadda, A., Stackpoole, M., and Stackpoole, M. (2019). Laser heat treatment improves quality in automotive die manufacturing. *Industrial Laser Solutions* 34 (1): 22–25.

Koshy, P. (1989). Laser cladding techniques for application to wear and corrosion resistant coatings. *SPIE, Applications of High Power Lasers* 527: 80–85.

Kou, S., Sun, D.K., and Le, Y.P. (1983). A fundamental study of laser transformation hardening. *Metallurgical Transactions A* 14A: 643–653.

Kusinski, J. and Thomas, G. (1986). Effect of laser hardening on microstructure and wear resistance in medium carbon/chromium steels. *SPIE, Laser Processing: Fundamentals, Applications, and Systems Engineering* 668: 150–157.

Lazare, S., Bolle, M., Cros, A., and Bellard, L. (1995). Periodic structuration of polymer surfaces with the excimer laser radiation. *Nuclear Instruments and Methods in Physics Research B* 105: 159–163.

Lee, D. and Kannatey-Asibu, E. Jr. (2011). Experimental investigation of laser shock peening using femtosecond laser pulses. *Journal of Laser Applications* 23 (2), 9 pages.

Li, W.-B., Easterling, K.E., and Ashby, M.F. (1986). Laser transformation hardening of steel – II. Hypereutectoid steels. *Acta Metallurgica* 34: 1533–1543.

Liao, Y., Suslov, S., Ye, C., and Cheng, G.J. (2012). The mechanisms of thermal engineered laser shock peening for enhanced fatigue performance. *Acta Materialia* 60: 4997–5009.

Lin, J. and Steen, W.M. (1998). Design characteristics and development of a nozzle for coaxial laser cladding. *Journal of Laser Applications* 10 (2): 55–63.

Maawad, E., Sano, Y., Wagner, L. et al. (2012). Investigation of laser shock peening effects on residual stress state and fatigue performance of titanium alloys. *Materials Science and Engineering A* 536: 82–91.

Matsunawa, A., Katayama, S., Minonishi, M. et al. (1991). Formation of alumina film by laser PVD. *ICALEO* 1991: 363–370.

Mazumder, J. (1983). Laser heat treatment: the state of the art. *Journal of Metals* 35: 18–26.

Montross, C.S., Wei, T., Ye, L. et al. (2002). Laser shock processing and its effects on microstructure and properties of metal alloys: a review. *International Journal of Fatigue* 24: 1021–1036.

Nalla, R.K., Altenberger, I., Noster, U. et al. (2003). On the influence of mechanical surface treatments/deep rolling and laser shock peening/on the fatigue behavior of Ti_6Al_4V at ambient and elevated temperatures. *Materials Science and Engineering* A355: 216–230.

Navara, E., Bengtsson, B., Li, W.B., and Easterling, K.E. (1984). Surface treatment of steel by laser transformation hardening. In: *International Congress on Heat Treatment of Materials: 07/11/1983-11/11/1983*, 40–44. Minerals, Metals & Materials Society.

Nerding, M., Christiansen, S., Dassow, R. et al. (2003). Tailoring texture in laser crystallization of silicon thin-films on glass. *Solid State Phenomena* 93: 173–178.

Peyre, P., Sollier, A., Chaieb, I. et al. (2003). FEM simulation of residual stresses induced by laser peening. *The European Physical Journal Applied Physics* 23 (2): 83–88.

Rankin, J., Hill, M., Halpin, J. et al. (2002). The effects of process variations on residual stress induced by laser peening. *Materials Science Forum* 404-407: 95–100.

Rouleau, B., Peyre, P., Breuils, J. et al. (2011). Characterization at a local scale of a laser-shock peened aluminum alloy surface. *Applied Surface Science* 257: 7195–7203.

Sano, T. (2018). Dry laser peening method uses femtosecond laser pulses. *Industrial Laser Solutions* 33 (4): 17–20.

Seaman, F.D. (1986). Laser heat-treating. In: *Industrial Laser Handbook*, 147–157. Tulsa: PennWell Books.

See, D.W., Dulaney, J., Clauer, A., and Tenaglia, R. (2002). The Air Force manufacturing technology laser peening initiative. *Surface Engineering* 18 (1): 32–36.

Semaltianos, N.G., Perrie, W., French, P. et al. (2008). Femtosecond laser surface texturing of a nickel-based superalloy. *Applied Surface Science* 255: 2796–2802.

Sepold, G. and Becker, R. (1986). Rapid solidification by laser beam techniques. In: *Science and Technology of the Undercooled Melt*, 112–119. Dordrecht: Springer.

Shewmon, P.G. (1983). *Atomic theory of diffusion, Diffusion in Solids*. Jenks, OK: J. Williams Book Company.

Shibata, K., Sakamoto, H., and Matsuyama, H. (1991). Absorptivity of polarized beam during laser hardening. *ICALEO* 1991: 409–413.

Shiue, R.K. and Chen, C. (1992). Laser transformation hardening of tempered 4340 steel. *Metallurgical Transactions A* 23A: 163–169.

Smith, P., Shepard, M., Prevey, P. III, and Clauer, A. (2000). Effect of power density and pulse repetition on laser shock peening of Ti–6Al–4V. *Journal of Materials Engineering and Performance* 9 (1): 33–37.

Solina, A., De Sanctis, M., Paganini, L. et al. (1984). Origin and development of residual stresses induced by laser surface-hardening treatments. *Journal of Heat Treating* 3: 193–204.

Tani, G., Orazi, L., Fortunato, A. et al. (2011). Warm laser shock peening: new developments and process optimization. *CIRP Annals - Manufacturing Technology* 60: 219–222.

Teresko, J. (2004). Peening with light: new laser, new benefits. *Industry Week* 253 (3): 20.

Van Brussel, B.A. and De Hosson, J.T.M. (1993). Residual stresses in the surface layer of laser-treated steels. *Materials Science and Engineering* A161: 83–89.

Vilhena, L.M., Sedlacek, M., Podgornik, B. et al. (2009). Surface texturing by pulsed Nd:YAG laser. *Tribology International* 42: 1496–1504.

Vincent, C., Monteil, G., Barriere, T., and Gelin, J.C. (2008). Control of the quality of laser surface texturing. *Microsyst Technologies* 14: 1553–1557.

Wang, A.A., Sircar, S., and Mazumder, J. (1990). Laser cladding of Mg–Al alloys. *ICALEO* 1990: 502–512.

Wei, M.Y. and Chen, C. (1993). Predicting case depth in tempered steels hardened via laser processing. *Materials Science and Technology* 9: 69–73.

Wu, X., Huang, C., Wang, X., and Song, H. (2011). A new effective method to estimate the effect of laser shock peening. *International Journal of Impact Engineering* 38: 322–329.

Ye, C., Suslov, S., Kim, B.J. et al. (2011). Fatigue performance improvement in AISI 4140 steel by dynamic strain aging and dynamic precipitation during warm laser shock peening. *Acta Materialia* 59: 1014–1025.

Yoshioka, Y., Akita, K., Suzuki, H. et al. (2002). Residual stress measurements of laser peened steels by using synchrotron radiation. *Materials Science Forum* 404–407: 83–88.

Zhang, W. and Yao, L. (2000). Improvement of laser induced residual stress distributions via shock waves. *ICALEO* 2000: 183–192.

Zhang, W., Yao, L., and Noyan, I. (2004a). Microscale laser shock peening of thin films, part 1: experiment, modeling and simulation. *ASME Journal of Manufacturing Science and Engineering* 126: 10–17.

Zhang, W., Yao, L., and Noyan, I. (2004b). Microscale laser shock peening of thin films, Part 2: high spatial resolution material characterization. *ASME Journal of Manufacturing Science and Engineering* 126: 18–24.

Zhou, Z., Gill, A.S., Qian, D. et al. (2011). A finite element study of thermal relaxation of residual stress in laser shock peened IN718 superalloy. *International Journal of Impact Engineering* 38: 590–596.

Zhou, Z., Bhamare, S., Ramakrishnan, G. et al. (2012). Thermal relaxation of residual stress in laser shock peened Ti–6Al–4V alloy. *Surface & Coatings Technology* 206: 4619–4627.

12 Laser Forming

Forming traditionally has been the manufacturing process by which the size or shape of a part is changed by the application of forces that produce stresses in the part that are greater than the yield strength and less than the fracture strength of the material. It requires the use of dies, which determine the final shape of the part. Simple examples of parts produced by this method are beverage containers, angle brackets, or connecting rods. Thermomechanical forming, on the other hand, enables parts (sheet metal, rod, pipe, or shell) to be formed without external forces and does not require the use of dies. Laser forming is a type of thermomechanical forming and may be used to form an angle bracket, for example, without using dies. More complex parts such as connecting rods that involve bulk forming, can only be made by traditional forming methods. However, where laser forming can be used, it also serves as a useful tool for rapid prototyping.

This chapter starts with a discussion on the basic principles of the laser-forming process. Typical process parameters that may be used are then outlined. This is followed by a discussion on the various mechanisms that underlie the process. The mechanics of the process are then discussed, followed by the advantages and disadvantages of the process. Finally, possible applications are presented.

12.1 PRINCIPLE OF LASER FORMING

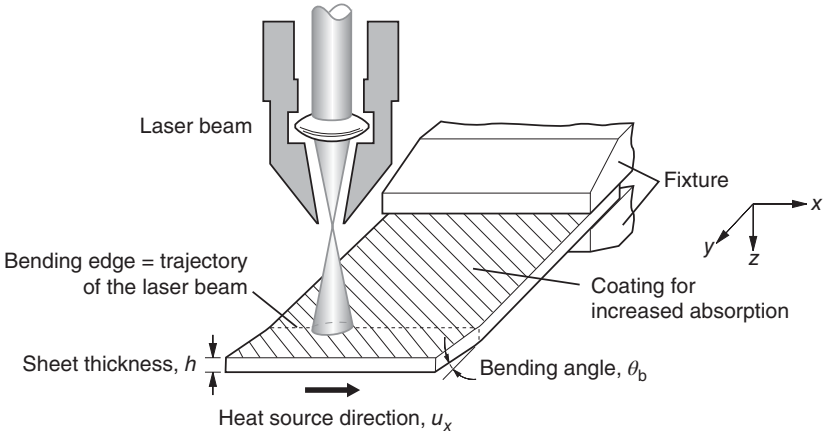
The technique for laser forming is very similar to that for laser surface heat treatment and involves scanning a defocused beam over the surface of the sheet metal to be formed. Moving the laser beam along a straight line without interruption causes the sheet to bend along the line of motion, Figure 12.1.

The components of a laser forming system include the following:

1. The laser source with beam delivery system.
2. Motion table unit on which the workpiece is mounted, or robot for holding a fiber-optic system.
3. Cooling system where necessary.
4. Temperature monitoring system.
5. Shape monitoring system.
6. Computer control system.

The typical bend angle that is achieved in a single step is about 2°, but may be as high as 10°. The total bend angle can be as high as 90° and higher by repetition of the process. The bend angle obtained after the first scan is greater than the bend angle obtained for each of the subsequent scans. However, after the first scan, the bend angle increases almost in proportion to the number of scans. More complex shapes can be obtained by offsetting each track by a small amount. In this

Principles of Laser Materials Processing: Developments and Applications, Second Edition. Elijah Kannatey-Asibu, Jr.
© 2023 John Wiley & Sons, Inc. Published 2023 by John Wiley & Sons, Inc.
Companion Website: www.wiley.com/go/Kannatey-Asibu/PrinicplesLaserMaterialsProcessing



(a)

(b)

Figure 12.1 (a) Schematic of the laser beam bending process. Source: From Vollertsen (1994b). (b) Photos of three sheet metals bent using a laser.

case, the radius of the part produced depends, among other processing parameters, on the amount of offset of each track. The smaller the offset, the smaller the radius.

The bend angle that is achieved for each step increases with a decrease in sheet thickness, Figure 12.2, due to a resulting decrease in bending restraint. The bend angle per step, however, decreases with a decrease in plate width. This is because with decreasing plate width, the volume of material that acts as a heat sink reduces, and as a result, the temperature gradient associated with the process decreases, resulting in a reduced compressive strain, and thus bend angle. On the other hand, for high width-to-thickness ratios greater than 10, the bend angle is almost independent of the plate width.

Processing parameters also affect the bend angle, with the angle increasing with increasing beam power and reducing scan rate. However, excessive power or very low scan rates will form a molten pool. This is undesirable since it will result in the surface coating being absorbed into the workpiece. Surface coating is normally used to increase the absorption of the laser radiation by the workpiece (Section 11.1.2.3). The method of cooling does not have very significant effect on the bend angle when laser forming.

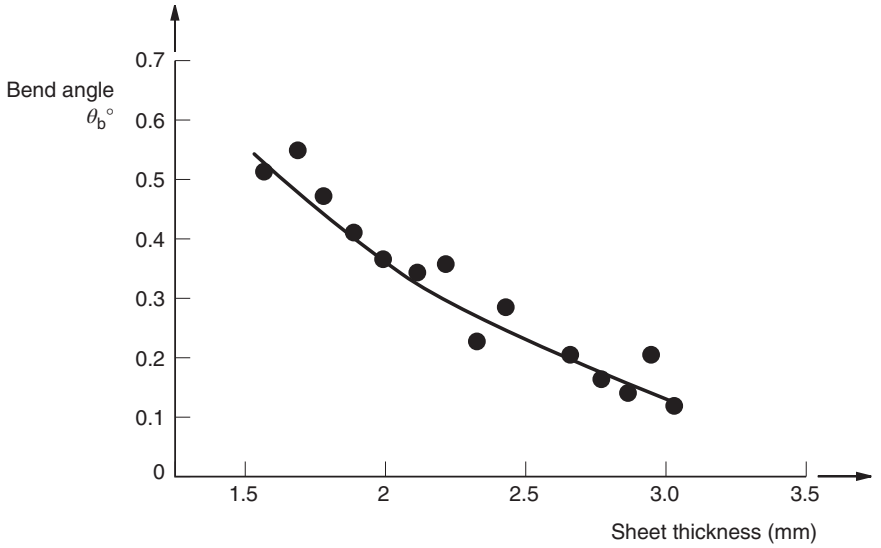


Figure 12.2 Influence of sheet thickness on bend angle for a plain carbon steel. Source: From Vollertsen et al. (1993)/Peter Lang Group AG.

12.2 PROCESS PARAMETERS

Typical processing conditions are listed in Table 12.1.

Thicker materials are more effectively bent by increasing the power input. However, the beam diameter needs to be increased accordingly to avoid melting the part. High scan rates are normally required for higher thermal conductivity materials such as aluminum alloys, while low thermal conductivity materials such as titanium alloys require lower scan rates. For the low thermal conductivity materials, if the scan rate is too high, only a thin surface layer will be heated. Efficiency of the process can be increased by proper application of absorptive coatings to increase the beam absorption, especially on metallic surfaces.

A major issue with the laser-forming process is the influence of the irradiation on the material's microstructure. Martensitic transformation, for example, may occur during laser forming of high carbon or alloy steels, and intermetallic phases may be formed in some aluminum alloys, resulting in corrosion.

12.3 LASER-FORMING MECHANISMS

The basic mechanism for laser forming is one of local deformation of the material as a result of thermally induced residual stresses. This may result from one of three process mechanisms depending on

Table 12.1 Typical process parameters used in laser forming.

Parameter	Value	Units
Power	1	kW
Defocused beam diameter	2–6	mm
Scanning velocities	20–600	mm/s
Sheet thickness	Up to 10	mm

the temperature field that is generated; the beam shape; and the dimensions of the part to be formed. The three mechanisms are

1. The temperature gradient mechanism.
2. The upsetting mechanism.
3. The buckling mechanism.

These are further discussed in the following sections.

12.3.1 Temperature Gradient Mechanism

The temperature gradient mechanism is illustrated in Figure 12.3. It evolves from processing conditions such as rapid heating of the workpiece surface, coupled with a relatively low heat conduction that generate very high temperature gradients in the thickness or z-direction, Figure 12.3a. Thermal expansion of the heated surface results in initial bending of the sheet away from the heat source, or toward the cold side of the workpiece during heating. This is often referred to as counterbending, and the bend angle at this stage is very small, $\approx 0.05°$. As the heated surface tries to expand, however, the constraint of the surrounding material, which is at a lower temperature, restricts its free expansion. This results in compressive thermal stresses that induce plastic compressive strain in the heated layer, Figure 12.3b. Plastic strains are induced since the yield strength of the material is lower at higher temperatures. With increasing ratio of plastic to elastic strain, the amount of final bend angle that results increases. Thus ideally, it is desirable to minimize the elastic strain by inducing very high temperatures in the top surface layers, where the flow stress is very low.

On cooling, the heated surface layer contracts more than the lower layers that are not heated as much as the surface layers, resulting in bending of the sheet toward the heat source, Figure 12.3c. Figure 12.4 illustrates a shape that is produced by the temperature gradient mechanism.

In the simplest case, a sheet of metal may be bent along a straight line. Repeating the process increases the bend angle, θ_b. During the bending process, the sheet thickness increases in the region along the bend as a result of the compressive strain. The temperature gradient mechanism is significant for relatively thick sheets. The ratio of the heated area diameter to the sheet thickness is relatively low, of the order of 1. It will not be effective for relatively thin materials of high thermal conductivity, since the resulting temperature gradient will then not be adequate to produce any effective bending.

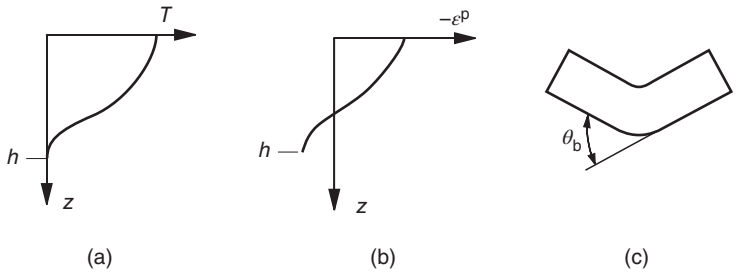


Figure 12.3 Temperature gradient mechanism. (a) Temperature variation in the thickness direction. (b) Strain variation in the thickness direction. (c) The bend angle. Source: Adapted from Geiger (1994).

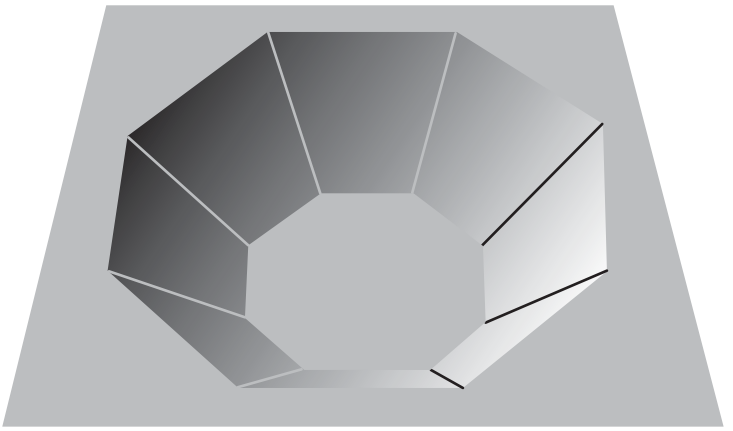


Figure 12.4 An illustration of a shape produced by the temperature gradient mechanism. Source: Adapted from Geiger (1994).

12.3.2 Buckling Mechanism

The buckling mechanism occurs in relatively thin sheets, where the ratio of the diameter of the heated area to the sheet thickness is relatively high, of the order of 10. The temperature gradient that arises in the thickness direction is then relatively small, so the material is heated almost uniformly through its thickness. The heated region tends to expand. However, this expansion is hindered by the unheated surrounding material. The restraint results in thermal compressive stresses being developed in the sheet, which for very thin sheets, may lead to buckling when a critical stress value is reached, Figure 12.5. Buckling is more likely to occur when the sheet is relatively thin. Otherwise, the critical load for buckling will not be reached.

Bending of material close to the center of the heat source occurs plastically due to the high temperatures in that region, while farther away, close to the root of the buckle, the material is subjected to elastic bending as a result of the lower temperatures in that region. If the heat source traverses the entire length of the sheet, then no restraining forces exist to hold the elastically bent portion in place during cooling. Thus, elastic recovery occurs at the root, while the center remains bent, resulting in the sheet being bent along the centerline. The direction of the bending that results from the buckling mechanism is unpredictable. It could either be toward or away from the heat source. This is because it is influenced by a number of factors such as the boundary conditions, the precurvature (or pre-bending) of the sheet, residual stresses, and external forces. However, it has been observed that at relatively high scan rates, say 15 mm/s (depending on the processing conditions), bending is always toward the laser beam. It only becomes unpredictable at lower speeds. Bending by the buckling mechanism does not increase the sheet thickness at the bend.

12.3.3 Upsetting Mechanism

The upsetting mechanism evolves when uniform heating of a localized zone is achieved through the thickness of the sheet, Figure 12.6a,b. Thus, the process parameters may be similar to those of the buckling mechanism, except for the diameter of the heat source area, which is relatively small. As a result of the near homogeneous heating of the sheet in the localized zone and prevention of

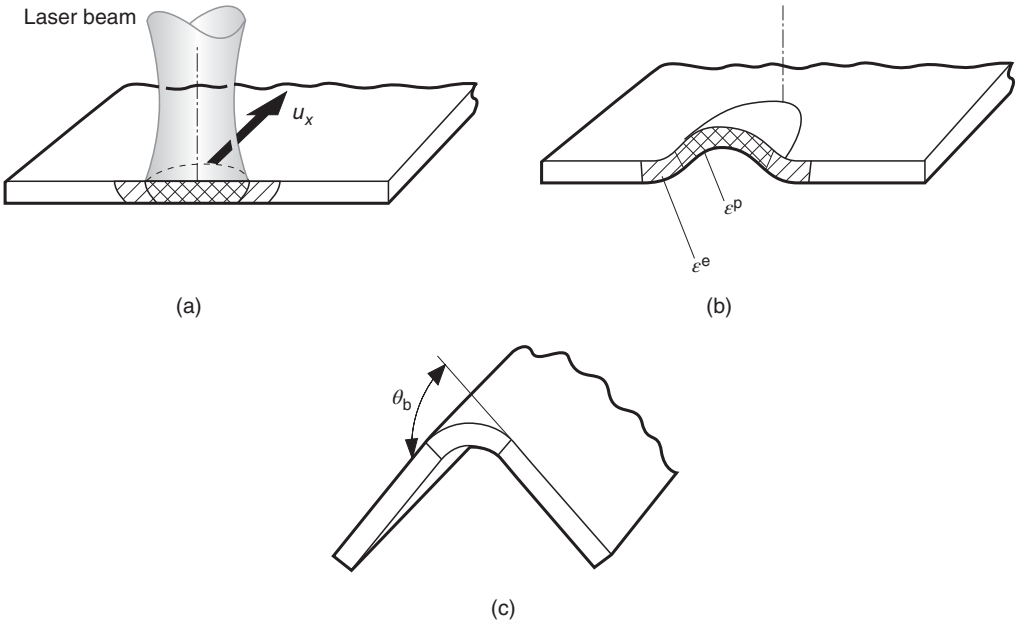


Figure 12.5 Buckling mechanism – sequence of steps leading to bending of a plate. (a) Beginning of heating. (b) Growth of buckle. (c) Development of bend angle. Source: From Vollertsen (1994a)/with permission of Meisenbach.

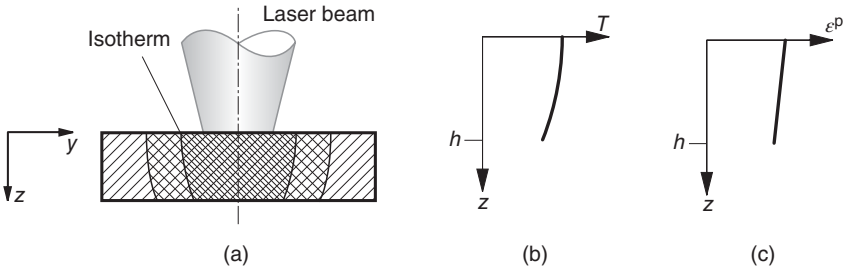


Figure 12.6 Upsetting mechanism. (a) Isotherms in the workpiece. (b) Temperature variation in the thickness direction. (c) Strain variation in the thickness direction. Source: From Vollertsen (1994a)/with permission of Meisenbach.

thermal expansion by the surrounding material, the sheet is subjected to near uniform compressive strain through its thickness, Figure 12.6c. During cooling, the heated region contracts, resulting in corresponding deformation of the sheet. This mechanism enables shapes that are similar to deep drawn parts to be produced, Figure 12.7.

12.3.4 Summary of the Forming Mechanisms

One common feature of all three mechanisms is the generation of strains by thermal stresses. For the temperature gradient and upsetting mechanisms, bending results primarily from plastic strains that

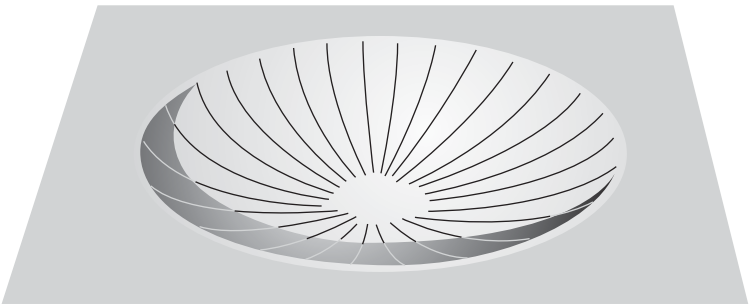


Figure 12.7 Sample product resulting from the upsetting mechanism. Source: From Vollertsen (1994a)/with permission of Meisenbach.

Table 12.2 Comparison of the three mechanisms for laser forming.

Parameter	Temperature gradient mechanism	Buckling mechanism	Upsetting mechanism
Temperature field	Steep temperature gradient	Nearly homogeneous with large diameter in x–y plane	Nearly homogeneous with small diameter in x–y plane
Sheet thickness	Increases	May remain constant	Increases
Forming direction	Toward laser beam; determined by process	Determined by precurvature, residual stresses	Mostly toward laser beam; depends on original part geometry
Applications	Bending sheets along straight lines	Bending sheets along straight lines. Tube forming	Spatial sheet metal forming; profile forming

Source: Vollertsen (1994a)/Meisenbach.

evolve during heating. Since elastic strains are reversible, their presence tend to reduce the bend angle, and are thus detrimental to the process. On the other hand, elastic strains are the driving force for forming in the case of the buckling mechanism.

The forming mechanisms are further summarized in Table 12.2.

12.4 PROCESS ANALYSIS

The laser forming process is a coupled problem that involves four basic concepts:

1. Analysis of the temperature field. This is essentially a three-dimensional conduction heat transfer problem involving a moving heat source. Approximate use may be made of the quasi-stationary solution discussed in Chapter 4. For a more accurate analysis, however, the time-dependent problem incorporating temperature-dependent material properties has to be solved using numerical methods, with the appropriate boundary conditions.
2. Analysis of the plastic strains using the appropriate constitutive laws, incorporating temperature-dependent thermal expansion coefficient, elastic modulus, and flow stress.

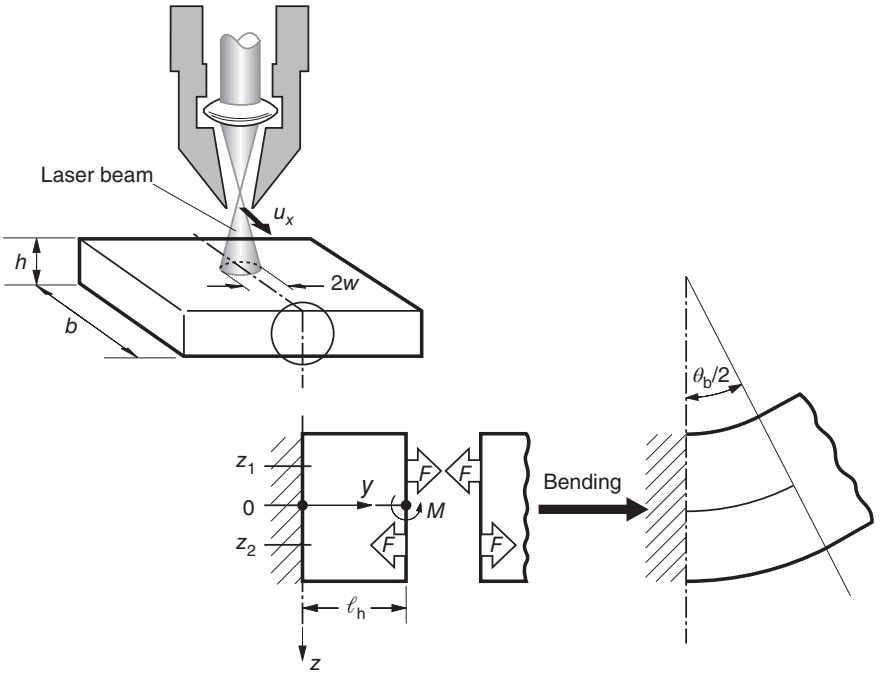


Figure 12.8 Two-layer model of laser beam bending. Source: From Vollertsen (1994b).

This would also involve a suitable yield criterion and restraint due to the geometry of the structure.

3. Analysis of the curvature during cooling due to the difference in plastic strains in the thickness direction. This would involve consideration of force and moment equilibrium, strain hardening, and flow stress.
4. Determination of the bend angle as an integral of the curvature that evolves with time.

In the remainder of this section, an approximate method is discussed. This provides some basic results that enable reasonable deductions to be made on the relationship between the process and product variables.

For the temperature gradient method, consider the plate shown in Figure 12.8, and an associated Cartesian coordinate system, *y*-*z*, with origin O directly below the centerline of the laser beam and halfway between the top and bottom surfaces. Assume that the sheet is made up of two layers in the heated zone of half length, l_h, and a single layer for $y > l_h$. Further assume that the plate width is relatively large compared to its thickness.

Let the centers of the two layers be located at z_1 and z_2. Then

$$z_1 - z_2 = \frac{h}{2} \tag{12.1}$$

where h is the thickness of the original sheet. For this simple case, the bend angle θ_b will be given by the geometry and difference in strains between the upper (ε_1) and lower (ε_2) layers:

$$\frac{\theta_b}{2} = \frac{l_h(\varepsilon_2 - \varepsilon_1)}{h/2} \tag{12.2}$$

where $\theta_b/2$ is the half bend angle that is based on the plate symmetry.

Let the lower layer be heated to a homogeneous temperature of T_l and the upper layer to a homogeneous temperature of $T_u = T_l + \Delta T$. During the heating phase, the upper layer will be subjected to compressive forces, while the lower layer is subjected to tensile forces. The strain, ε_1, to which the upper layer is subjected will be a superposition of the strains due to the induced force, F, moment, M, and thermal expansion:

$$\varepsilon_1 = \frac{F}{E_{m1}A_1} - \frac{M_1}{(E_m I_m)_1} z_1 + \beta_l \Delta T \tag{12.3}$$

where A is the cross-sectional area (m^2), E_m is the elastic modulus (N/m^2), I_m is the second moment of area (m^4), M is the moment due to the forces (N m), F is the force acting on each layer (N), β_l is the linear coefficient of thermal expansion (/K), and ΔT is the temperature rise of the upper layer relative to the lower layer (K). The subscripts 1 and 2 indicate the upper and lower layers, respectively.

This analysis is based on the assumption that the thermal expansion of the upper layer is entirely converted into compressive plastic strain. The assumption neglects the temperature dependence of the yield stress, as well as the reversed straining during cooling. This is reasonable since at high temperatures, the yield stress of the material is relatively low.

As the sheet cools, conduction of heat into the lower layer results in a tendency for it to expand. This creates compressive stresses in this layer and that reduces the amount of tensile stress that develops in the upper layer. Thus, there is no cancellation of the compressive strain in the lower layer. The conversion of thermal expansion into plastic compression then results in

$$\varepsilon_1 = \frac{F}{E_{m1}A_1} - \frac{M_1}{(E_m I_m)_1} z_1 - \beta_l \Delta T_m \tag{12.4}$$

The corresponding strain in the lower layer is

$$\varepsilon_2 = -\frac{F}{E_{m2}A_2} - \frac{M_2}{(E_m I_m)_2} z_2 \tag{12.5}$$

where ΔT_m is the maximum temperature difference between the upper and lower layers.

The local curvature is given by the ratio of the moment M to the section modulus $E_m I_m$. For large curvatures, it can be assumed that

$$\frac{\theta_b}{2l_h} = \frac{M_1}{(E_m I_m)_1} = \frac{M_2}{(E_m I_m)_2} = \frac{M}{E_m I_m} \tag{12.6}$$

Now, the bending moment is given by $M = F \times h/2$. Combining this with Eq. (12.6) gives

$$F = \frac{2M}{h} = \frac{\theta_b E_m I_m}{h l_h} \tag{12.7}$$

From the preceding equations and considering that when $h_1 = h_2 = h/2$ and $I_m = bh^3/12$ for a rectangular plate that is heated through the upper half, we get

$$\theta_b = \frac{3 l_h \beta_l \Delta T_m}{h} \tag{12.8}$$

And using an approximate solution for the temperature field, Eq. (12.8) can also be expressed as

$$\theta_b = \frac{3\beta_l q}{\rho c_p u_x h^2} \tag{12.9}$$

where q is the power absorbed by the surface. The bend angle calculated is in radians. For the general case where the thicknesses of the upper and lower layers are different, then we have

$$A_1 = bh_1 \tag{12.10}$$

and

$$A_2 = b(h - h_1) \tag{12.11}$$

The bend angle then becomes

$$\theta_b = \frac{12\beta_l \Delta T_m l_h h_1 (h - h_1)}{h^3} \tag{12.12}$$

An approximate expression for estimating the bend angle that results for each pass of the laser beam during laser forming has been derived in this section. The analysis is based on the temperature gradient mechanism. Similar expressions may be obtained for the other mechanisms. However, only one approach is presented since they are all approximates. The interested reader is referred to (Vollertsen, 1994a, b). More exact determination of the deformed shape will require setting up the appropriate governing equations for both the heat flow and deformation and solving them numerically.

Example 12.1 A titanium plate is bent using a 1.5 kW Nd:YAG laser. The plate dimensions are thickness 4.5 mm, width 200 mm, and length 600 mm. The beam radius is 4 mm, and scanning is done at a rate of 40 mm/s along the width of the sheet metal. For a single scan, determine

1. The bend angle, θ_b.
2. The induced force, F.
3. The bending moment, M.
4. Strains, ε_1 and ε_2, in the upper and lower parts of the sheet, respectively.

Assume a surface reflection coefficient of 85%.

Solution:
From appendices 4D and 4E, the properties of the titanium plate are approximated as average density, $\rho = 4507\,\text{kg/m}^3$; average specific heat, $c_p = 99.3\,\text{J/kg K}$; thermal conductivity, $k = 0.0114\,\text{W/mm K}$; modulus of elasticity, $E = 116\,\text{GPa}$ (Appendix 7.B); linear coefficient of thermal expansion, $\beta_l = 10.8 \times 10^{-6}/\text{K}$.

1. From Eq. (12.9), the bend angle for a single pass is given by

$$\theta_b = \frac{3\beta_l q}{\rho c_p u_x h^2} = \frac{3 \times 10.8 \times 10^{-6} \times (1 - 0.85) \times 1500}{4.507 \times 10^{-6} \times 99.3 \times 40 \times 4.5^2} = 0.02\,\text{rad} = 1.15^\circ$$

2. The induced force is obtained from Eq. (12.7) as

$$F = \frac{2M}{h} = \frac{\theta_b E_m I_m}{h l_h}$$

The second moment of area is given by

$$I_{\mathrm{m}} = \frac{bh^3}{12} = \frac{200 \times 4.5^3}{12} = 1518.75 \ \ \mathrm{mm}^4$$

Now if the half-length of the heated zone is estimated as being equal to the beam radius, then $l_{\mathrm{h}} = 4$ mm. Thus,

$$F = \frac{0.02 \times 116 \times 10^3 \times 1518.75}{4.5 \times 4} = 196 \ \ \mathrm{kN}$$

3. The bending moment is obtained from Eq. (12.7) as

$$M = F \times \frac{h}{2} = 196{,}000 \times \frac{4.5}{2} = 440 \times 10^3 \ \mathrm{N\,mm} = 440 \ \mathrm{N\,m}$$

4. The strains are given by Eqs. (12.4) and (12.5) as

$$\varepsilon_1 = \frac{F}{E_{\mathrm{m}1}A_1} - \frac{M_1}{(E_{\mathrm{m}}I_{\mathrm{m}})_1}z_1 - \beta_1 \Delta T$$

$$\varepsilon_2 = -\frac{F}{E_{\mathrm{m}2}A_2} - \frac{M_2}{(E_{\mathrm{m}}I_{\mathrm{m}})_2}z_2$$

The maximum temperature difference between the upper and lower layers can be obtained from Eq. (12.8) as

$$\Delta T_{\mathrm{m}} = \frac{h\theta_{\mathrm{b}}}{3l_{\mathrm{h}}\beta_{\mathrm{l}}} = \frac{4.5 \times 0.020}{3 \times 10.8 \times 10^{-6} \times 4} = 694 \ \ ^\circ\mathrm{C}$$

Therefore,

$$\varepsilon_1 = \frac{196 \times 10^3}{116 \times 10^3 \times 450} - \frac{440 \times 10^3}{(116 \times 10^3 \times 1518.75)_1} \times 1.125 - 10.8 \times 10^{-6} \times 694 = -0.00655$$

and

$$\varepsilon_2 = -\frac{196 \times 10^3}{116 \times 10^3 \times 450} - \frac{440 \times 10^3}{(116 \times 10^3 \times 1518.75)_1} \times 1.125 = -0.00657$$

12.5 ADVANTAGES AND DISADVANTAGES

12.5.1 Advantages

The primary advantages of laser forming include

1. The noncontact nature of the process, which makes it independent of tool inaccuracies that might result from wear and deflection, since no external forces are involved. It also makes the process more flexible.
2. Ability to more accurately control the energy source and thus the forming process, compared to flame bending and mechanical forming in general.
3. Forming of parts in confined or inaccessible locations by remote application.
4. Minimal heat-affected zone size or material degradation compared to flame bending where the heat source is more diffuse.
5. Ability to form hard and brittle materials that are difficult to form by conventional methods.

12.5.2 Disadvantages

The major limitations are

1. The process is relatively slow compared to mechanical forming, by a factor of 5–20.
2. The high capital cost of the laser makes it more expensive than other forms of thermomechanical forming such as flame bending.
3. The low efficiency of the laser results in high energy consumption.
4. Due to the potential reflection of the laser beam in various directions, special safety precautions are essential.

12.6 APPLICATIONS

One area in which laser forming finds useful application is part straightening for mass production. This involves using the laser beam to deform and thus straighten parts that become distorted at one stage or another of their production. Due to the small size of the laser beam, and the ease with which it can be controlled, laser forming is highly suited for forming small components, that is, micrometal forming. Laser forming is also useful for forming of some automobile body components, even though it may not be suitable for the high production rates essential for automobile components. On the other hand, due to the low-volume requirements in forming of ship planks and production of aerospace fuselage, laser forming is found to be useful for such applications. Also, due to the high flexibility of the laser beam, it is well suited to the production of sheet metal components in space. Finally, laser forming has been applied to metal foams in recent years. Metal foams have high strength to weight ratios and are excellent for shock and noise absorption. They are therefore useful in the aerospace and automotive industries. The temperature gradient mechanism is the primary mechanism encountered with metal foams. One challenge with metal foams is localized melting of thin cell walls.

Since laser forming is slow relative to conventional forming, it is most economical for maintenance work when there are no spare parts available. Also, since it does not require dies, laser forming is suited for complex components in small-lot production and fabrication of individual parts for rapid prototyping. This enables the performance of several variants of a design to be evaluated before committing to full production equipment.

12.7 SUMMARY

Laser forming is a thermomechanical forming process, and thus enables sheet metal parts and pipes to be formed without external forces. Scanning the laser beam along a straight line causes the sheet metal to bend along the line of motion. More complex parts, however, can be formed by using a more complex trajectory. The three principal process mechanisms are

1. The temperature gradient mechanism – which results from processing conditions that generate very high-temperature gradients in the sheet metal.
2. The buckling mechanism – which is more common in relatively thin sheets, where the ratio of the diameter of the heated area to the sheet thickness is relatively high of the order of 10.
3. The upsetting mechanism – which evolves when uniform heating of a localized zone is achieved through the thickness of the sheet, using a heat source area that is relatively small.

Analysis shows that the bend angle achieved is directly proportional to the power input and inversely proportional to the traverse velocity, as well as the inverse of the plate thickness squared.

Compared to traditional sheet metal forming, laser forming is relatively slow. However, it is noncontact and is also suitable for applications such as rapid prototyping and micrometal forming. In addition, compared to other thermomechanical forming processes, it is appropriate for situations where very little microstructural damage to the base material can be sustained.

APPENDIX 12.A

List of symbols used in the chapter

Symbol	Parameter	Units
I_m	Second moment of area	m^4
l_h	Half length of heated layer	mm
T_l	Temperature of lower layer	°C
T_u	Temperature of upper layer	°C
ΔT	Temperature difference between upper and lower layers	°C
ΔT_m	Maximum temperature difference between upper and lower layers	°C
z_1, z_2	Centers of two layers in the heated zone	mm
$\varepsilon_1, \varepsilon_2$	Strains in upper and lower layers	—
θ_b	Bend angle	rad

PROBLEMS

12.1. A low carbon steel sheet metal is bent to form a shallow V-shape using a 2 kW CO_2 laser. The beam radius is 5 mm, and scanning is done at a rate of 25 mm/s along the width of the sheet metal. If the sheet thickness is 3 mm, width 250 mm, and length 400 mm, determine, after the first pass

1. The bend angle.
2. The induced force, F.

Assume a surface reflection coefficient of 75%.

12.2. For the conditions in problem 12.1, determine

1. The bending moment, M, that is induced during the process.
2. The strains, ε_1 and ε_2, in the upper and lower parts of the sheet, respectively.
3. How much time will be required to bend the sheet metal to an angle of 45°, if the beam rapid traverse is neglected?

12.3. A 6061-T6 aluminum alloy of thickness 4 mm, length 250 mm, and width 50 mm is to be bent along its width to form a 90° angle bracket for an automobile component. If the desired production rate for the part is 100 pieces per hour,

1. Calculate the power input to the workpiece that will be required for the process.
2. Select an appropriate laser for the process and estimate its desired power output.

12.4. A titanium plate is to be bent slightly to an angle of 2.5° using a laser. The plate width is 75 mm and length 250 mm. How much power will be required to complete the process in a single scan if the beam radius is 5 mm, scanning is done at a rate of 50 mm/s along the width of the sheet metal, and the temperature rise between the top and bottom layers is $\Delta T_m = 775\,°C$?

Assume a surface reflection coefficient of 75%. What force will be induced in the workpiece under these conditions?

12.5. A 2024 aluminum alloy is formed using a laser of output power 4 kW at a scan rate of 42 mm/s, with the beam diameter being 6 mm. The reflection coefficient is 30%. If the sheet thickness is 4 mm and the width of the rectangular workpiece in the direction of motion of the heat source is 20 mm, with the length dimension being 50 mm, determine

1. The bend angle.
2. The bending moment that is induced during the process.

12.6. A sheet of steel is bent by passing a laser over the bend line. However, in addition to the moment that results due to the laser, M_l, an external moment, M_e, is also applied to the sheet, in the same direction as the bend moment due to the laser. Find an equation for the bend angle, accounting for the addition of the external moment.

BIBLIOGRAPHY

Bao, J. and Yao, L. (1999). *Analysis and Prediction of Edge Effects in Laser Bending*, 253–262. San Diego, CA: *ICALEO*.

Bucher, T., Cardenas, S., Verma, R. et al. (2018). Laser forming of sandwich panels with metal foam core. *ASME Journal of Manufacturing Science and Engineering* 140 (11): 111015–111015-12. https://doi.org/10.1115/1.4040959.

Chen, G. and Xu, X. (2002). Experimental and 3D finite element studies of CW laser forming of thin stainless steel sheets. *ASME Journal of Manufacturing Science and Engineering* 123: 66–73.

Frackiewicz, H. (1993). Laser metal forming technology. *Fabtech, International'93, The Society of Manufacturing Engineers*, pp. 733-747.

Geiger, M. (1994). Synergy of laser material processing and metal forming. *Annals of the CIRP* 43: 1–8.

Geiger, M., Vollertsen, F., and Deinzer, G. (1993). *Flexible straightening of car body shells by laser forming*, 37–44. Detroit, Michigan: *SAE International Congress and Exposition*.

Magee, J., Watkins, K.G., Steen, W.M. et al. (1998). Laser bending of high strength alloys. *Journal of Laser Applications* 10 (4): 149–155.

Masubuchi, K. (1992). Studies at M.I.T. related to applications of laser technologies to metal fabrication. *Proceedings of LAMP* '92: 939–946.

Namba, Y. (1986). Laser forming in space. In: *International Conference on Lasers* (ed. C.P. Wang), 403–407.

Namba, Y. (1987). Laser forming of metals and alloys. *Proceedings of LAMP'87* , pp. 601-606.

Pridham, M.S. and Thomson, G.A. (1994). Laser forming: a force for the future? *Materials World* 2: 574–575.

Santo, L., Bellisario, D., Rovatti, L., and Quadrini, F. (2012). Microstructural modification of laser-bent open-cell aluminum foams. *Key Engingeering Materials* 504–506: 1213–1218.

Scully, K. (1987). Laser line heating. *Journal of Ship Production* 3: 237–246.

Thomson, G.A. and Pridham, M.S. (1995). Laser forming. *Manufacturing Engineer* 74: 137–139.

Vaccari, J.A. (1993). The promise of laser forming. *American Machinist* 137: 36–38.

Veiko, Y.P. and Yakovlev, Y.B. (1994). Physical fundamentals of laser forming of micro optical components. *Optical Engineering* 33: 3567–3571.

Vollertsen, F., Geiger, M., and Li, W.M. (1993). FDM and FEM simulation of laser forming: a comparative study. *Advanced Technology of Plasticity 1993 – Proceeding of the Fourth International Conference on Technology of Plasticity*, pp. 1793–1798.

Vollertsen, F. (1994a). Mechanisms and models for laser forming. *Laser Assisted Net shape Engineering Proceedings of the LANE '94* 1: 345–360.

Vollertsen, F. (1994b). An analytical model for laser bending. *Lasers in Engineering* 2: 261–276.

Vollertsen, F., Komel, I., and Kals, R. (1995). The laser bending of steel foils for microparts by the buckling mechanism – A model. *Modeling and Simulations in Materials Sciences and Engineering* 3: 107–119.

Zhang, W., Yao, L., and Noyan, I. (2004a). Microscale laser shock peening of thin films, part 1: experiment, modeling and simulation. *ASME Journal of Manufacturing Science and Engineering* 126: 10–17.

Zhang, W., Yao, L., and Noyan, I. (2004b). Microscale laser shock peening of thin films, part 2: hhigh spatial resolution material characterization. *ASME Journal of Manufacturing Science and Engineering* 126: 18–24.

Zhang, X.R., Chen, G., and Xu, X. (2002). Numerical simulation of pulsed laser bending. *ASME Journal of Applied Mechanics* 69: 254–260.

Zhang, X.R. and Xu, X. (2003). High precision microscale bending by pulsed and CW lasers. *ASME Journal of Manufacturing Science and Engineering* 125: 512–518.

13 Additive Manufacturing

Additive manufacturing (AM), in its basic form, involves the generation of physical objects from graphical computer data without the use of traditional tools. It is referred to in the literature by several other names, for example, three-dimensional (3D) printing, rapid prototyping, stereolithography, solid freeform manufacturing (fabrication), layer manufacturing, desktop manufacturing, and direct digital manufacturing. In its early years, it was used mainly in design conceptualization to facilitate engineering stress analysis and in making patterns and/or dies for manufacturing operations. However, the technology has improved to the point that it is now used extensively for end products that include aerospace and automotive components. It is most suitable for custom products, especially complex parts that are difficult to produce by traditional methods. In that regard, as the technology evolves, it will eventually put a significant amount of manufacturing in the hands of the consumer, taking us closer to the concept of personalized production.

We start this chapter with a general discussion on the basic concepts of additive manufacturing, namely, computer-aided design (CAD), part building (processing), and post-processing. The major types of additive manufacturing systems in current use (including non-laser-based systems) are outlined under part building. In this respect, this chapter is different from other chapters, since it also includes processes that are not laser-based. This is followed by a discussion on the main applications of additive manufacturing. A general overview of additive manufacturing is illustrated in Figure 13.1, indicating some of the major processes currently in use.

Producing a part by additive manufacturing typically involves three main stages:

1. CAD.
2. Part building (processing).
3. Post-processing.

These basic steps are common among the different types of additive manufacturing techniques, and will be discussed in that order. The specific additive manufacturing techniques will be discussed under part building.

13.1 COMPUTER-AIDED DESIGN

The first stage of additive manufacturing involves either using an advanced 3D CAD system to develop a computer model of the part to be produced or using reverse engineering to develop a computer model from a part. The reverse engineering approach normally involves using, for example, either a coordinate measuring machine (CMM) or a laser digitizer to trace out the part. However, a reasonably good CAD model can now be obtained using the camera of an adapted cellular phone.

Principles of Laser Materials Processing: Developments and Applications, Second Edition. Elijah Kannatey-Asibu, Jr.
© 2023 John Wiley & Sons, Inc. Published 2023 by John Wiley & Sons, Inc.
Companion Website: www.wiley.com/go/Kannatey-Asibu/PrinicplesLaserMaterialsProcessing

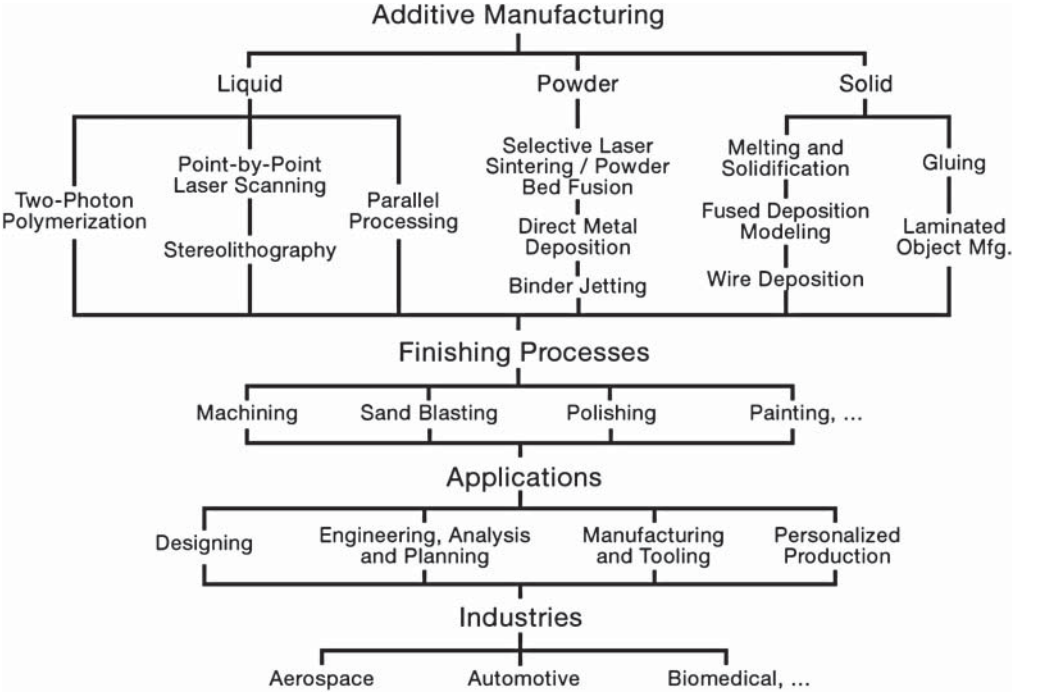


Figure 13.1 General overview of additive manufacturing systems.

For the computer model to be most useful, it needs to be represented as closed surfaces that define an enclosed volume, thus ensuring that all horizontal cross sections are closed curves. A number of techniques are available for representing computer models of objects, and they often make use of geometric transformations. These techniques normally involve the creation of curves, surfaces, and solids. The latter is generally referred to as solid modeling and is most suited for additive manufacturing applications. We briefly discuss curves, surfaces, and solids in that order.

13.1.1 Curve and Surface Design

Curve and surface design is important since all objects are bounded by curves and surfaces. They are especially useful in the design of aircraft fuselage, automobile bodies, ship hulls, and sculptures. Two common curve forms that are used in geometric modeling are splines and Bezier curves.

13.1.1.1 Splines A spline is essentially a smooth curve that is made to go through a set of points. The simplest form is the cubic spline which can be used to fit a given set of data points, $x_0, x_1, \ldots, x_n$. This is done by first fitting each contiguous pair of points, (x_j, x_{j+1}), with a cubic polynomial such that the polynomial and its first and second derivatives are continuous on a set of intervals from x_0 to x_n.

A cubic polynomial can generally be expressed as:

$$y = f(x) = \sum_{i=0}^{3} a_i x^i \tag{13.1}$$

Thus, the cubic spline function can be expressed in the following form:

$$S_j(x) = a_{0j} + a_{1j}x + a_{2j}x^2 + a_{3j}x^3, \qquad j = 0, 1, 2, \ldots, n-1 \tag{13.2}$$

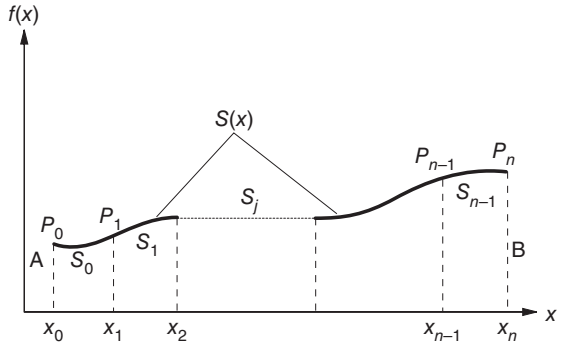


Figure 13.2 A spline function.

where S_j represents span j of the spline, with the following conditions:

$$S(x_j) = f(x_j), \qquad j = 0, 1, 2, \ldots, n \tag{13.3}$$

$$S_{j+1}(x_{j+1}) = S_j(x_{j+1}), \qquad j = 0, 1, 2, \ldots, n-2 \tag{13.4}$$

$$S'_{j+1}(x_{j+1}) = S'_j(x_{j+1}), \qquad j = 0, 1, 2, \ldots, n-2 \tag{13.5}$$

$$S''_{j+1}(x_{j+1}) = S''_j(x_{j+1}), \qquad j = 0, 1, 2, \ldots, n-2 \tag{13.6}$$

The spline function is fully determined when the coefficients, a_{ij}, in Eq. (13.2) are determined. This is done by applying Eqs. (13.3)–(13.6) to each span (S_j) of the spline, that is, between two contiguous points, (x_j, x_{j+1}). Unfortunately, this will result in a system of $n-1$ equations with $n+1$ unknowns. The other two equations are obtained by considering the boundary conditions at the end points of the spline, A and B (Figure 13.2). If these are given by

$$S''(x_0) = 0 \tag{13.7a}$$

$$S''(x_n) = 0 \tag{13.7b}$$

then the resulting spline is a natural spline. It is called a clamped spline if the boundary conditions are specified at the ends as follows:

$$S'(x_0) = f'(x_0)$$

$$S'(x_n) = f'(x_n)$$

where $f'(x_0)$ and $f'(x_n)$ are specified.

Equation (13.2) defines a nonparametric cubic spline, and these have the property that they are single-valued functions. This makes them unsuitable for multi-valued or closed curves, which are often encountered in geometric modeling. This problem may be overcome with the use of parametric splines. In parametric form, the variables x and y, Eq. (13.1), are expressed in terms of a third variable, t, such that

$$x = x(t) \tag{13.8a}$$

$$y = y(t) \tag{13.8b}$$

The cubic spline between two points is then expressed as:

$$S(t) = \sum_{i=0}^{3} a_i t^i \quad t \in [0, 1] \tag{13.9}$$

And a point on the curve is given by:

$$S(t) = [x(t) \quad y(t)] \tag{13.10}$$

Again, the constants a_i in Eq. (13.9) are obtained by considering the boundary conditions for each spline and then also the overall boundary conditions for a natural or clamped spline.

13.1.1.2 Bezier Curves The major setback of cubic splines is that the curves are uniquely defined within a given interval, and thus the shape cannot be controlled. Bezier curves permit control of the curve's shape. They are not constrained to pass through all the specified points. They only have to pass through the end points. The data points define the vertices of a polygon which defines the shape of the curve. The curve is tangent to the polygon at the end points and lies within the convex of the polygon, as shown in Figure 13.3. The mathematical basis of the Bezier curve is derived from the Bernstein polynomial, which is given, for a degree n, by:

$$B_{n,i}(t) = \begin{bmatrix} n \\ i \end{bmatrix} t^i (1-t)^{n-i}, \quad i = 0, 1, \ldots, n \tag{13.11}$$

where

$$\begin{bmatrix} n \\ i \end{bmatrix} = \frac{n!}{i!(n-i)!}$$

$n!$ is the n factorial, given by $n! = n\,(n-1)\,(n-2)\,\ldots\,2{\cdot}1$. The Bezier curve is given by:

$$S(t) = \sum_{i=0}^{3} P_i B_{n,i}(t) \quad t \in [0, 1] \tag{13.12}$$

where $S(t)$ denotes points on the curve, P_i indicates specified data points, t is selected increment, n is degree of the Bezier polynomial, and $n+1$ is number of data points, giving $n+1$ vertices of the polygon.

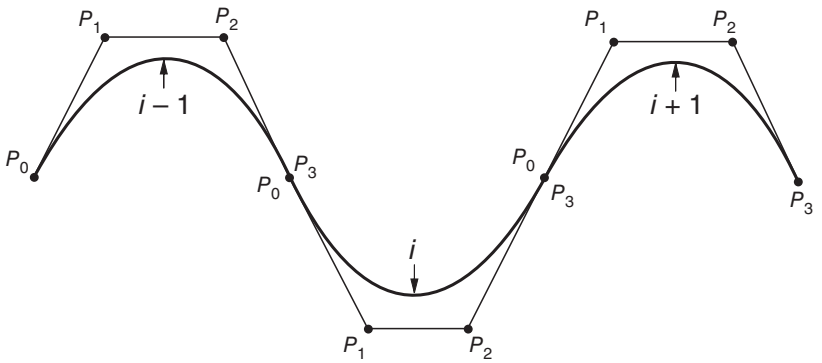


Figure 13.3 A Bezier curve.

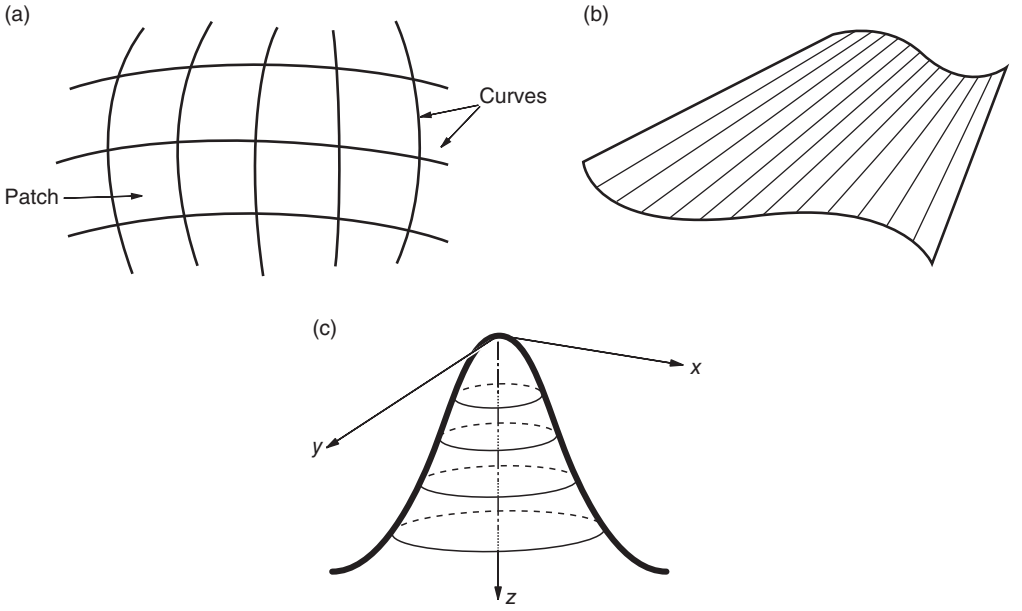


Figure 13.4 (a) A surface constructed from patches: (b) a ruled surface and (c) a surface of revolution.

For a 3D curve, we have

$S(t) = [x(t)\ y(t)\ z(t)]$ = points on the curve.
$P_i = [x_i\ y_i\ z_i]$ = coordinates of the data points.

13.1.1.3 Surface Representation A common approach used in constructing complex surfaces is to concatenate several small patches, as shown in Figure 13.4a. Another type is the ruled surface which contains two space parametric curves as opposite boundary curves. Linear interpolation is then used to connect the two curves at several points, thereby defining a surface, as shown in Figure 13.4b. A curve may also be rotated about an axis to generate a surface of revolution, as shown in Figure 13.4c. A patch, for example, may be generated as a Bezier surface, which, for a region bounded by four curves, is expressed by:

$$S(s,t) = \sum_{i=0}^{m} \sum_{j=0}^{n} P_{i,j} G_{i,m}(s) G_{j,n}(t) \tag{13.13}$$

where $P_{i,j}$ indicates the specified data points $(m{+}1\ \ n{+}1)$, $G_{i,m}(s) = \frac{m!s^i(1-s)^{m-i}}{i!(m-i)!}$, and $G_{j,n}(t) = \frac{n!t^j(1-t)^{n-j}}{j!(n-j)!}$.

A point on the surface is given by $S(s, t) = [x(s, t)\ y(s, t)\ z(s, t)]$.

Example 13.1 Generate the Bezier curve for the data points given by P0 [1, 2], P1 [2, 3], P2 [3, 0], and P3 [4, 1]

Solution:
Since there are four points, $n = 3$. From Eq. (13.11), we obtain $B_{n,i}(t)$ for $n = 3$, $i = 0$ to 3 as follows:

$B_{3,0}(t) = (1-t)^3$

Table 13.1 Bezier function values for different t values.

t	$B_{3,0}(t)$	$B_{3,1}(t)$	$B_{3,2}(t)$	$B_{3,3}(t)$
0	1	0	0	0
0.25	0.421	0.421	0.14	0.015
0.5	0.125	0.375	0.375	0.125
0.75	0.015	0.14	0.421	0.421
1	0	0	0	1

$B_{3,1}(t) = 3t(1-t)^2$
$B_{3,2}(t) = 3t^2(1-t)$
$B_{3,3}(t) = t^3$

With these functions, the Bezier curve is now given by:

$$S(t) = P_0B_{3,0} + P_1B_{3,1} + P_2B_{3,2} + P_3B_{3,3}$$

For values of t = 0.0, 0.25, 0.5, 0.75, 1.0, the corresponding Bezier coefficients are listed in Table 13.1.

Points on the Bezier curve are then given by:

$S(0) = [1, 2]\cdot 1 + [2, 3]\cdot 0 + [3, 0]\cdot 0 + [4, 1]\cdot 0 = [1, 2]$
$S(0.25) = [1, 2]\cdot 0.421 + [2, 3]\cdot 0.421 + [3, 0]\cdot 0.14 + [4, 1]\cdot 0.015 = [1.743, 2.12]$
$S(0.5) = [1, 2]\cdot 0.125 + [2, 3]\cdot 0.375 + [3, 0]\cdot 0.375 + [4, 1]\cdot 0.125 = [2.5, 1.5]$
$S(0.75) = [1, 2]\cdot 0.015 + [2, 3]\cdot 0.14 + [3, 0]\cdot 0.421 + [4, 1]\cdot 0.421 = [3.242, 0.871$
$S(1) = [1, 2]\cdot 0 + [2, 3]\cdot 0 + [3, 0]\cdot 0 + [4, 1]\cdot 1 = [4, 1]$

The corresponding Bezier curve is shown in the following graph:

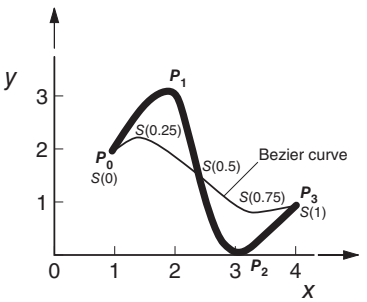


Figure E13.1

CAD systems are now primarily based on the NURBS (nonuniform rational basis splines) representation.

13.1.2 Solid Modeling

The curve generation methods discussed in Section 13.1.1 are useful for generating wireframe models of objects. The wireframe method has the advantage of simplicity and speed of implementation.

Even though quite useful for visualizing certain types of objects, it has a number of disadvantages. Since hidden lines cannot be removed, representation of complex objects using the wireframe method can often be ambiguous. Furthermore, it does not incorporate information on the physical characteristics of the object such as volume and mass. These drawbacks are eliminated with the use of solid models. Two of the most common solid modeling methods are constructive solid geometry (CSG) and boundary representation (BREP).

13.1.2.1 Constructive Solid Geometry (CSG) The CSG technique builds 3D objects using primitives, which are basic geometric shapes such as cubes, cylinders, cones, and spheres as building blocks. Boolean operators such as union (∪), intersection (∩), and difference (−) are used to combine them to create the final object. Figure 13.5 illustrates these operations using simple geometric figures. The process of union combines primitives into a new object, while the process of difference may be used to create a hole, for example. Intersection results in an object whose volume is common to the initial primitives and is useful for checking interference between objects.

The objects formed at intermediate stages of the CSG product development are organized in an orderly manner using a tree structure. A tree node represents an intermediate solid, which is obtained by combining primitive and intermediate solids lower down in the tree's hierarchy. This is illustrated for a simple object in Figure 13.6.

One advantage of the CSG method is that it does not require as much memory storage as the other solid modeling methods, even though the computational time for constructing the object may be more. Also, the structure of the CSG tree generally parallels the sequence of operations that will be required to manufacture the object. Thus, the designer is able to envisage the basic manufacturing plan for the object.

13.1.2.2 Boundary Representation (BREP) With the BREP method, objects are defined by their enclosing surfaces or boundaries, in terms of their faces, edges, and vertices. Boolean operations may then be used to combine these entities to form the desired object. The end result is similar to a wireframe object but can incorporate the object's physical properties. The BREP technique is useful for nonstandard shapes such as aircraft fuselage and automobile bodies and enables models to be constructed faster. However, it requires more storage, but less computational time in constructing the object.

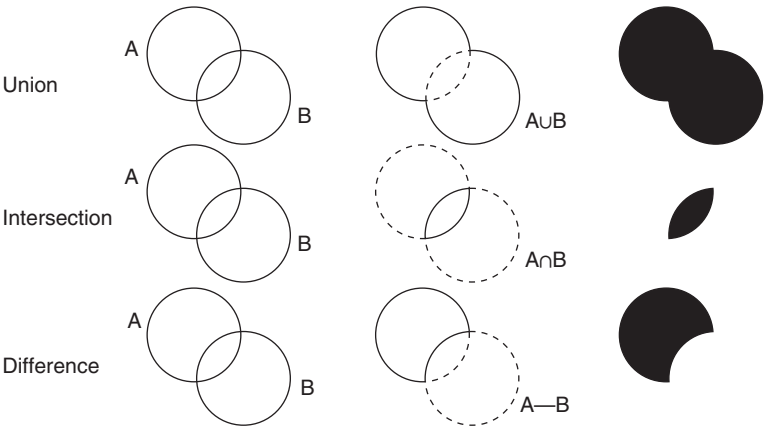


Figure 13.5 Boolean operations on simple geometric figures.

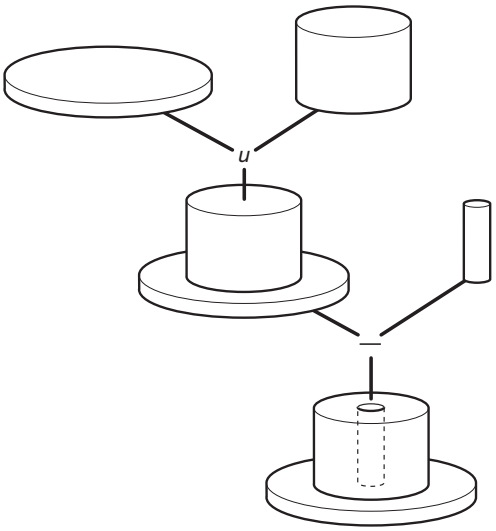


Figure 13.6 Tree structure with Boolean operations for a simple object.

A number of systems use a combination of these basic concepts. Common CAD systems that are based on solid modeling include Solid Works, I-DEAS, Unigraphics II, Pro/Engineer, CATIA, and Alias Designer. Generally, for convenience, and also to avoid potential errors downstream, all CAD values are positive numbers. That means the part geometry must reside in the positive x, y, z octant.

13.1.3 Software Formats

The CAD model is next converted into a format required by the additive manufacturing (AM) equipment. After the conversion, the file is then transmitted to the AM system's computer. Two of the formats used are the standard tessellation language (STL) and initial graphics exchange specifications (IGES) formats.

13.1.3.1 The STL Format This was developed by 3D Systems and is the most commonly used format. It approximates the outside surfaces of the model using polygons, usually, triangular facets. For example, a rectangle is represented using two triangles, while a curved surface is approximated using a number of triangles. Of course, for curved surfaces, the larger the number of triangles used, the more accurate the representation. However, the file size or computation time then increases accordingly, even though the subsequent build (processing) time is not significantly affected. The desired accuracy is achieved by specifying a part resolution. This may be expressed either as a maximum absolute deviation from the true surface (Figure 13.7a) or as a percentage of chord length (Figure 13.7b) depending on the CAD system.

Each triangular facet is described by a set of X, Y, and Z coordinates for each of the three vertices and a unit normal vector to indicate which side of the facet is inside the object. Since STL files do not contain topological data, the polygonal approximation models used tend to exhibit a number of errors including gaps and overlapping facets (resulting from numerical round-off errors).

The STL format has the advantage that it is simple, providing simple files for data transfer. However, it has the disadvantage that its file is much larger than the original CAD file, since it carries a lot of redundant information. In addition, it is susceptible to the geometric errors mentioned in the preceding paragraph. Slicing of the files takes a relatively long time due to the large file sizes.

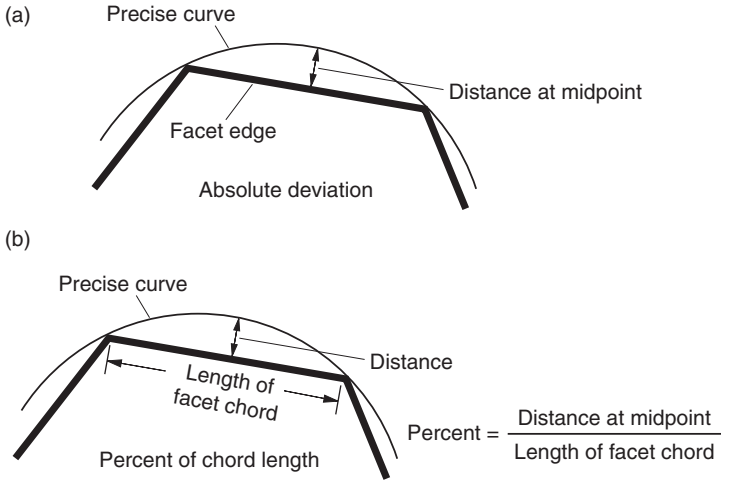


Figure 13.7 Accuracy specification for a CAD system. (a) Absolute deviation. (b) Percent of chord length. Source: Jacobs (1992)/with permission of SME.

13.1.3.2 The IGES Format This is a standard used to exchange graphics information between commercial CAD systems. It is capable of precisely representing CAD models and includes not only geometry information but also topological information. The entities of points, lines, arcs, splines, surfaces, and solid elements are provided. Ways of representing the regularized operations for union, intersection, and difference are defined. IGES files, however, have the disadvantage that they often contain redundant information that is not needed for AM systems. Furthermore, the algorithms for slicing (Section 13.1.5) an IGES file are more complex than those for slicing an STL file. AM systems often need support structures (Section 13.1.4). These cannot be created directly using the IGES format.

13.1.4 Supports for Part Building

In additive manufacturing, supports are often needed to hold the part in place during the building phase (Figure 13.8). They also allow for safe removal of the part from the platform and prevent the recoater from colliding with the platform during the build stage. The supports are created in the CAD system, but generally, the output files are separate from those of the part. Creating special supports is essential for most of the liquid-based systems (Section 13.2.1), while for other systems such as laminated object manufacturing, selective laser sintering (SLS), and binder jetting, no special supports are necessary since the supports are inherent in the part buildup. Where needed, supports may be generated by the engineer, CAD designer, or AM operator.

Sections for which supports are critical include areas that overhang preceding layers. If such a section is not supported, it will tend to pull on the preceding layer as it shrinks, resulting in curl, just as occurs in a bimetallic strip. Supports normally consist of combinations of thin webs, and they need to overlap into the succeeding part layers by two to three layer thicknesses. An example of a support system is shown in Figure 13.8.

13.1.5 Slicing

The additive manufacturing process generally involves building the part layer by layer, one on top of the other. To facilitate this process, the CAD model is first sliced into cross sections (Figure 13.9).

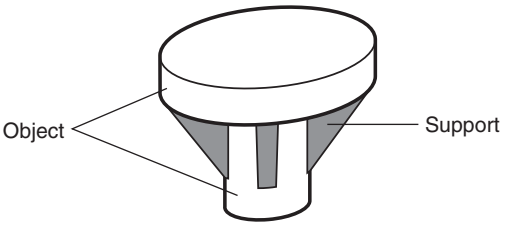


Figure 13.8 Sample support structure.

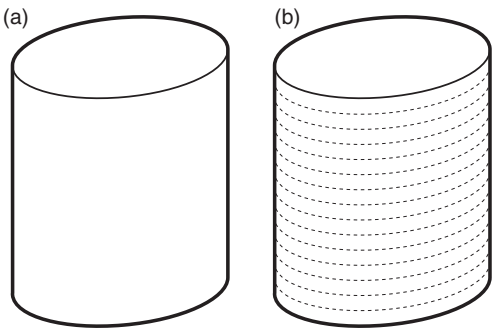


Figure 13.9 Slicing of a CAD model. (a) Original object. (b) Sliced object.

The slice thickness selected depends on the accuracy desired, and it typically ranges from 0.12 to 0.50 mm. Variation of the build time with slice thickness is shown in Figure 13.10 for liquid-based systems. For very thin layers, the build time decreases as layer thickness increases. However, at relatively large layer thicknesses, the build time begins to increase with increasing layer thickness. This is because the scan velocity decreases exponentially with increasing cure thickness.

The CAD stage is normally the most time-consuming part of the additive manufacturing process. CAD systems that are used are normally those which have capability for other design applications such as interference studies, stress analysis, FEM analysis, detail design and drafting, and planning of manufacturing (including NC programming). This ensures that the same CAD package can be used for other design applications. In the next section, we discuss the basic concepts associated with part building.

13.2 PART BUILDING

There are several types of additive manufacturing methods, and these can be categorized by the configuration of the light source used, the materials, the energy source, or the type of application. Due to the significant changes in processing methodology associated with different materials, our discussion on part building will be based on the type of material to be processed. In this regard, we consider three main types of systems:

1. Liquid-based systems.
2. Powder-based systems.
3. Solid-based systems.

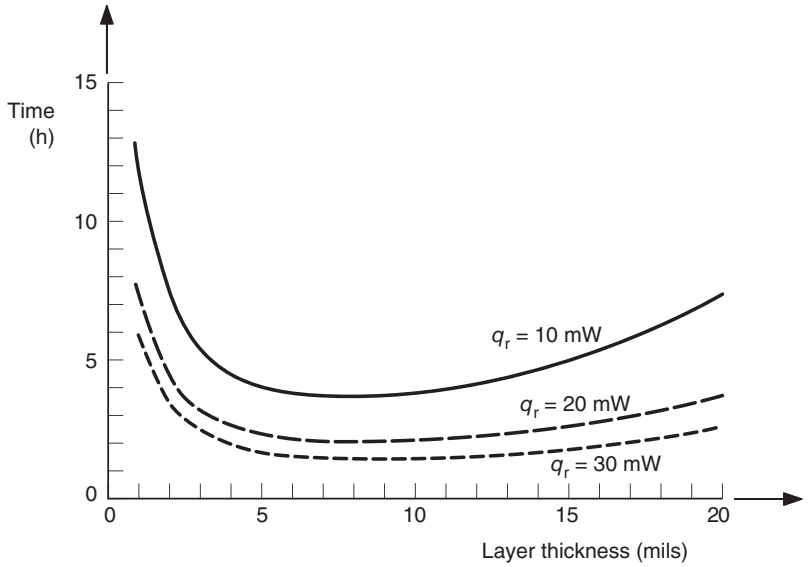


Figure 13.10 Variation of build time with slice thickness for liquid-based systems. Source: Jacobs (1992)/with permission of SME.

13.2.1 Liquid-Based Systems

Liquid-based additive manufacturing is normally referred to as stereolithography (SL) or vat polymerization. The basic premise of the liquid-based systems involves building parts in a vat of photocurable organic liquid resin (usually thermosets) by exposing it to laser radiation to catalyze polymerization by inducing cross-linking of the photopolymer and thus solidify or cure it. This involves linear absorption of the radiation by the organic molecules. The wavelength of the laser radiation used is relatively short with high photon energy and mainly in the ultraviolet range. Thus, the most commonly used lasers are He-Cd and argon ion lasers. At any one time, only a thin layer (with depth equal to the slice thickness) of the resin is exposed to adequate enough radiation for polymerization to occur. The principle is illustrated in Figure 13.11.

Before each layer is exposed to the laser beam, it needs to be levelled. This involves first checking the resin's z level (height) using a sensor. If the resin level falls outside a tolerance band, a plunger is used to reposition the part in the z-direction until the desired level is achieved. Following levelling, the platform (on which the part rests) is lowered (deep dipping). A few seconds (z-wait), typically about 15–30 seconds, is then allowed for the resin to settle. The resin is now exposed to the laser beam. When the layer has solidified, the part is lowered to permit the next layer of resin to be formed over it. This is repeated until the part is completed. This is the general process, even though details may vary from one manufacturer to another.

A major advantage of stereolithography is that it is more accurate than many of the other additive manufacturing techniques listed. Even though almost all the processes enable internal hollow spaces to be produced, stereolithography has the added advantage that it enables the hollow interior to be completely drained. However, the process has the disadvantage that it is only useful for photosensitive materials. Furthermore, the end product needs to have UV protective coating; otherwise, it is subject to rapid degradation. Finally, it is a two-step process involving first the laser treatment, followed by curing in an oven.

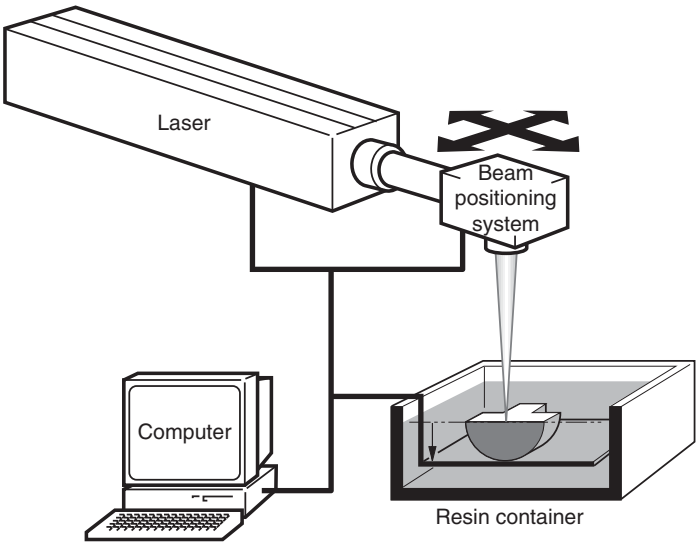


Figure 13.11 Basic principle of stereolithography. Source: Kochan (1993)/with permission of Elsevier.

There are three basic ways in which the resin can be exposed to the laser beam:

1. One-dimensional (1D) scanning with
   a. A single source (beam scanning).
   b. Array of sources.
2. Two-dimensional (2D) source (parallel processing).
3. Two-photon polymerization.

These techniques for energy or material deposition are also applicable to other additive manufacturing methods and are now discussed in Sections 13.2.1.1–13.2.1.3.

13.2.1.1 Beam Scanning In beam scanning, the laser beam traces a path on the resin surface that is determined by the shape to be formed at that level (Figure 13.11). Scanning can be done using either a galvanometer, robot, or *X*–*Y* motion unit. First, the part borders for the given layer cross-section are traced. The areas that will eventually become solid are then filled in or hatched. This essentially involves scanning the interior. There are different hatching styles available (Figure 13.12a,b). The scan lines of a given layer may be either perpendicular to each other or a combination of horizontal and 60°/120° lines which produce an internal pattern consisting of equilateral triangles. Scan lines of contiguous slices (layers) may either overlap (Figure 13.12c) or just touch each other (Figure 13.12d). They may also be in line (Figure 13.12d) or be staggered (Figure 13.12e). Hatch spacings are typically about 0.25 mm. Hatching results in vertical columns containing unsolidified resin, which is usually cured during post-processing. Staggering and overlapping the scan lines has the advantage of eliminating microfissures that tend to occur with the simple hatch without overlap or staggering.

Sample conditions for additive manufacturing by beam scanning are shown in Table 13.2.

For effective scanning, the depth to which the laser beam penetrates the liquid has to at least equal the layer thickness. That defines the depth profile of the hatch lines shown in Figure 13.12. To estimate the depth profile, let us consider the coordinate system as shown in Figure 13.13. The intensity

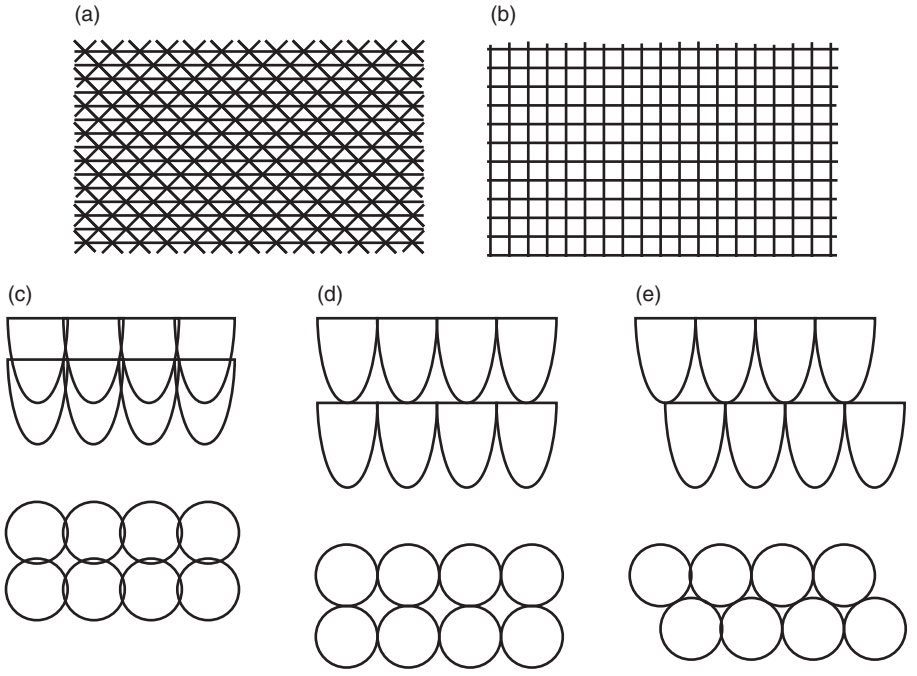


Figure 13.12 Hatch types. (a) Triangular hatch. (b) Rectangular hatch. (c) Overlapping layers. (d) Low overlapping layers. (e) Staggered hatch. Source: Kochan (1993)/with permission of Elsevier.

Table 13.2 Sample conditions for stereolithography by beam scanning.

Process parameter	Range of values	Units
Laser power	5–500	mW
Beam size	50–300	μm
Scan speed	0.75–24	m/s
Minimum layer thickness	0.01–0.3	mm

Source: From Chua and Leong (1997).

of the beam propagation can then be expressed as:

$$I(z) = I_0 e^{-\frac{z}{D_p}} \tag{13.14}$$

where I_0 is the intensity of the incident beam (W/cm^2) and $D_p = 1/\alpha$ = penetration depth (mm).

For a Gaussian beam, the distribution of the incident beam, I_0, on the surface of the liquid layer is given by:

$$I_0(r) = I_p e^{\left(-\frac{2r^2}{w^2}\right)} \tag{13.15}$$

where $I_0(r)$ is the beam intensity at any radial position, r, on the surface (W/cm^2), I_p is the peak beam intensity (W/cm^2), r is radial coordinate (μm), and w is the radius at which beam intensity is $1/e^2 \times I_p$ or $0.14 \times I_p$. That means 86% of the total energy is contained in a spot of radius w (μm).

The peak intensity in Eq. (13.15), I_p, can be shown to be given (Problem 13.4) by:

$$I_p = \frac{2q}{\pi w^2} \tag{13.16}$$

Thus, the surface intensity (power per unit area) becomes

$$I_0(r) = \frac{2q}{\pi w^2} e^{\left(-\frac{2r^2}{w^2}\right)} \tag{13.17}$$

Since the laser beam is considered to scan linearly in the x-direction, by integrating Eq. (13.17) over the period of scanning (mathematically from $x = -\infty$ to $x = \infty$, but in reality the time it takes the laser beam to traverse its practical zone of influence), we get the exposure (energy per unit area) on the surface, Q_0, (Problem 13.5) as:

$$Q_0(y) = \sqrt{\frac{2}{\pi}} \frac{q}{w u_x} e^{\left(-\frac{2y^2}{w^2}\right)} \tag{13.18}$$

Equation (13.18) is a function only of the y coordinate, since we are considering the exposure on the surface ($z = 0$), and quasi-stationary conditions are assumed, that is, there is no change in the x-direction. Now taking into account the propagation of the beam in the depth or z-direction using Eq. (13.14), we get the exposure through the thickness of the slice as:

$$Q_p(y, z) = \sqrt{\frac{2}{\pi}} \frac{q}{w u_x} e^{\left(-\frac{2y^2}{w^2}\right)} e^{-\frac{z}{D_p}} \tag{13.19}$$

For a given level of exposure, the profile of the region that is affected by the laser beam is as shown in Figure 13.13. Thus if the resin that is being treated has a critical exposure value, Q_c, above which it polymerizes, then we can determine the cross section (y–z or depth boundary) of the region

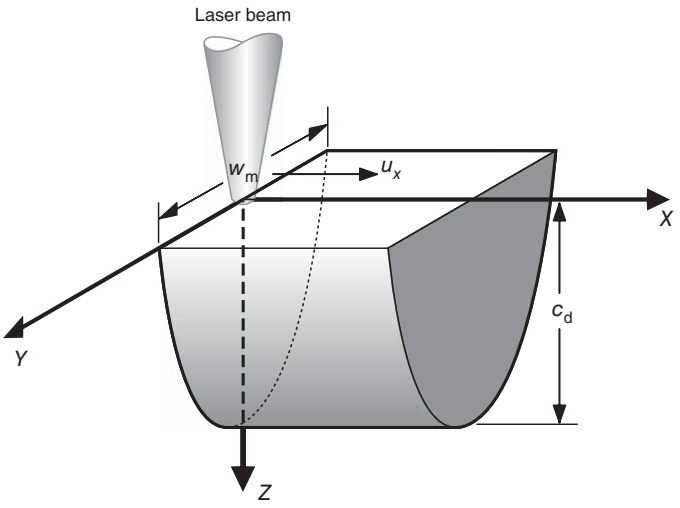


Figure 13.13 Depth profile of a laser scan. Source: Adapted from Jacobs (1992).

that is polymerized by the laser beam for each scan, by setting $Q_p = Q_c$, that is,

$$Q_c = \sqrt{\frac{2}{\pi}} \frac{q}{w u_x} e^{\left(-\frac{2y_c^2}{w^2}\right)} e^{-\frac{z_c}{D_p}} \tag{13.20}$$

where y_c, z_c = values of y, z when $Q_p = Q_c$.

Equation (13.20) can be shown (Problem 13.6) to reduce to the form:

$$z_c = a y_c^2 + b \tag{13.21}$$

where $a = -\frac{2D_p}{w^2}$ and $b = D_p \ln\left[\sqrt{\frac{2}{\pi}}\left(\frac{q}{w u_x Q_c}\right)\right]$.

This is the equation of a parabola. Thus, the scanned line has a parabolic cross section (Figure 13.13). From Figure 13.13, it is evident that the maximum penetration that is cured (cure depth), C_d, can be obtained by setting y_c to zero. That gives

$$C_d = b = D_p \ln\left[\sqrt{\frac{2}{\pi}}\left(\frac{q}{w u_x Q_c}\right)\right] \tag{13.22}$$

Furthermore, the maximum width of the cured region (which occurs on the slice surface), W_m, is given by:

$$W_m = w\sqrt{2\frac{C_d}{D_p}} \tag{13.23}$$

Example 13.2 An argon ion laser of output power 260 mW is used to process a photocurable resin at a scan speed of 2 m/s, with a beam radius of 200 μm. If the resin has a critical exposure of 12 mJ/cm^2 and the depth, z_c, at a distance of $y_c = 0.15$ mm from the scan line on the surface is 0.4 mm, determine the cure depth, C_d, for the process.

Solution:
From Eq. (13.20), we have

$$Q_c = \sqrt{\frac{2}{\pi}} \frac{q}{w u_x} e^{\left(-\frac{2y_c^2}{w^2}\right)} e^{-\frac{z_c}{D_p}}$$

$$\Rightarrow 12 \times 10^{-5} = \sqrt{\frac{2}{\pi}} \frac{260 \times 10^{-3}}{200 \times 10^{-3} \times 2000} e^{\left(-\frac{2\times 0.15^2}{200^2 \times 10^{-6}}\right)} e^{-\frac{0.4}{D_p}}$$

Thus,

$$12 \times 10^{-5} = 16.84 \times 10^{-5} e^{-\frac{0.4}{D_p}}$$

$$\Rightarrow D_p = 1.18 \text{ mm}$$

Now from Eq. (13.22), the cure depth is obtained as:

$$C_d = b = D_p \ln\left[\sqrt{\frac{2}{\pi}}\left(\frac{q}{w u_x Q_c}\right)\right]$$

$$\Rightarrow C_d = 1.18 \times \ln\left[\sqrt{\frac{2}{\pi}}\left(\frac{260 \times 10^{-3}}{200 \times 10^{-3} \times 2000 \times 12 \times 10^{-5}}\right)\right] = 1.72 \text{ mm}$$

The most common liquid-based 1D systems include

1. Stereolithography by 3D Systems.
2. Stereography by EOS.

13.2.1.2 Parallel Processing With parallel processing, a mask is used to expose the entire area of each layer to be solidified to the laser beam at the same time (Figure 13.14). A different mask is used for each layer, and the exposed area of the mask corresponds to the area to be solidified at that layer. This process is thus faster than the beam scanning process. It is also cheaper since the same glass plate is used to generate each of the masks. Several techniques have been developed for masking, but one simple approach involves transferring the information for each layer to a glass plate by electrostatically charging the areas to be masked, such that toner adheres to those areas. The regions corresponding to the part to be produced for that layer are thus left transparent, enabling the ultraviolet laser radiation to be transferred to the resin, polymerizing it in those regions. The unexposed regions remain liquid. After each layer is exposed to the beam, the resin that is not polymerized is sucked away and replaced by water soluble wax, which provides support for the part. Thus, no supports are needed.

Parallel processing systems are produced by 3D Systems and EnvisionTEC.

13.2.1.3 Two-Photon Polymerization Two-photon polymerization (2PP) is a process that is used to fabricate 3D micro-components that have nano-features, using photosensitive resins, and is based on the two-photon absorption phenomenon. It enables 3D polymer-based objects to be created within bulk resin material without resorting to layer-by-layer application of the resin, as is done in conventional additive manufacturing.

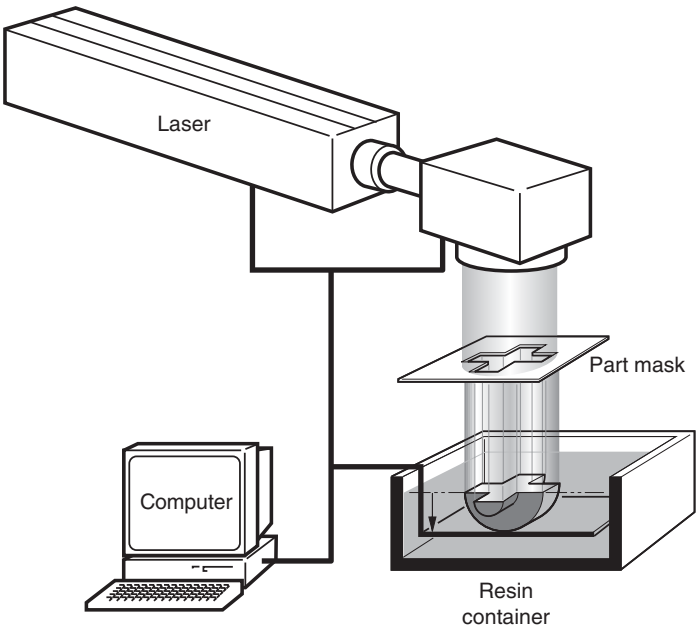


Figure 13.14 Parallel exposure of the resin to the laser beam.

The polymerization process is photoinitiated (as opposed to thermal initiation), and thus by controlling the intensity of the beam, the generation of radicals can be controlled. As a result, the process can be controlled with high precision. The resin used in photopolymerization is normally one that is catalyzed by ultraviolet radiation, together with a photochemical initiator which absorbs the radiation and becomes decomposed into free radicals. In 2PP, an infrared wavelength laser beam is used so that instead of the single ultraviolet photon, two photons of near-infrared wavelength (each of half the energy required) are simultaneously absorbed.

In addition to the photoinitiators, the resin also consists of monomers and oligomers, together with an appropriate concentration of reaction terminators. An example would be urethane acrylate monomers/oligomers. The process of radical polymerization involves initiation, propagation, and termination and is illustrated in Eqs. (13.24)–(13.27). In photopolymerization, in general, one quantum of light decomposes a single initiator molecule into two free radicals. Thus, each initiator (I), which normally absorbs a single ultraviolet photon, absorbs two infrared photons simultaneously and is decomposed into two free radicals ($2\dot{R}$), as shown in Eq. (13.24). Each radical reacts with a monomer (M) or oligomer (M_n), generating a chain radical ($R\dot{M}$) at the ends of the monomers and oligomers, as shown in Eq. (13.25). Each new radical combines with another monomer, generating a chain reaction, as shown in Eq. (13.26), which is terminated when the chained radical meets another chained radical, as shown in Eq. (13.27).

Initiation:

$$\mathrm{I} + h_p\nu + h_p\nu \rightarrow 2\dot{\mathrm{R}} \tag{13.23}$$

$$\dot{\mathrm{R}} + \mathrm{M} \rightarrow \mathrm{R}\dot{\mathrm{M}} \tag{13.25a}$$

For example,

$$\dot{\mathrm{R}} + \mathrm{CH_2{=}CHX} \rightarrow \mathrm{RC\dot{H}_2C}\begin{matrix}\mathrm{H}\\|\\\mathrm{X}\end{matrix} \tag{13.25b}$$

Propagation:

$$\mathrm{R}\dot{\mathrm{M}} + \mathrm{M} \rightarrow \mathrm{R}\dot{\mathrm{M}}_1 \tag{13.26a}$$

$$\mathrm{R}\dot{\mathrm{M}}_n + \mathrm{M} \rightarrow \mathrm{R}\dot{\mathrm{M}}_{n+1} \tag{13.26b}$$

For example,

$$\mathrm{R\text{-}(CH_2CHX\text{-})_n C\dot{H}_2}\begin{matrix}\mathrm{H}\\|\\\mathrm{C}\\|\\\mathrm{X}\end{matrix} + \mathrm{CH_2{=}CHX} \rightarrow \mathrm{R-(CH_2CHX\text{-})_{n+1}\,C\dot{H}_2}\begin{matrix}\mathrm{H}\\|\\\mathrm{C}\\|\\\mathrm{X}\end{matrix} \tag{13.26c}$$

Termination by combination:

$$\mathrm{R}\dot{\mathrm{M}}_n + \mathrm{R}\dot{\mathrm{M}}_n \rightarrow \mathrm{RM}_{2n}\mathrm{R} \tag{13.27a}$$

For example,

$$-\mathrm{C\dot{H}_2}\begin{matrix}\mathrm{H}\\|\\\mathrm{C}\\|\\\mathrm{X}\end{matrix} + \begin{matrix}\mathrm{H}\\|\\\mathrm{C}\\|\\\mathrm{X}\end{matrix}\mathrm{\dot{C}H_2}- \rightarrow -\mathrm{CH_2}\begin{matrix}\mathrm{H}\\|\\\mathrm{C}\\|\\\mathrm{X}\end{matrix}-\begin{matrix}\mathrm{H}\\|\\\mathrm{C}\\|\\\mathrm{X}\end{matrix}\mathrm{CH_2}- \tag{13.27b}$$

Or termination by disproportionation:

$$R\dot{M}_n + R\dot{M}_n \rightarrow RM_n + RM_n \tag{13.27c}$$

For example,

$$\begin{array}{c} H \quad H \\ | \quad | \\ -C\dot{H}_2\,C + C\dot{C}H_2- \end{array} \rightarrow \begin{array}{c} H \qquad H \\ | \qquad | \\ -CH_2\,C - H + C = CH- \end{array} \tag{13.27d}$$
$$\begin{array}{c} | \quad | \\ X \quad X \end{array} \qquad\qquad \begin{array}{c} | \qquad | \\ X \qquad X \end{array}$$

Here, R is a photoinitiator and results from the removal of an electron from $\dot{R}$. Addition of a monomer to an oligomer M_n through the propagation process results in the new oligomer M_{n+1}. An oligomer consists of a finite number of monomers, as opposed to a polymer which theoretically has an infinite number of monomers.

The main reason for using an infrared wavelength laser is that since the resin is transparent to the infrared radiation, the beam is able to propagate, with little or no attenuation, to the focal zone. Material in the unfocused region (both solid and liquid) is not affected by the radiation. If a femtosecond laser of high enough intensity is used with tight focusing, the power density in the focal zone can initiate two-photon absorption in that region. Since the rate of two-photon absorption is proportional to the square of the power density, the process is limited to a small region within the focal zone. Moving the focal point in 3D space through the resin enables any computer-generated 3D structure to be fabricated within the interior of the resin (Figure 13.15), with a resolution better than 100 nm, and there is no need to fabricate it layer by layer, as is done with conventional additive manufacturing. This improves the resolution of the process. The tight focusing also enables spatial resolution smaller than the laser diffraction limit to be achieved. If ultraviolet radiation was used (in which case the process would be single photon), the beam would polymerize material (while also being attenuated by it) through which it propagates.

As an example, the liquid resin ORMOCER, which is an organic–inorganic hybrid polymer, is sensitive to the 390 nm radiation and if it is exposed to a tightly focused femtosecond laser radiating at 780 nm (Ti : sapphire) with a pulse duration of 80 fs, and a pulse rate of 80 MHz, 2PP occurs. In the setup shown in Figure 13.16, a microscope objective with a 100× magnification and a high

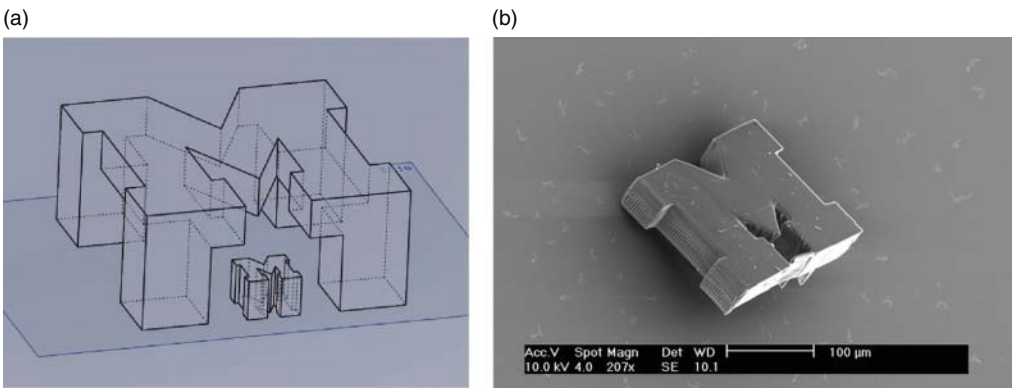


Figure 13.15 3D object produced by two-photon polymerization. (a) Original CAD representation. (b) Microscale object produced by 2PP. Source: Gregory (2012)/University of Michigan.

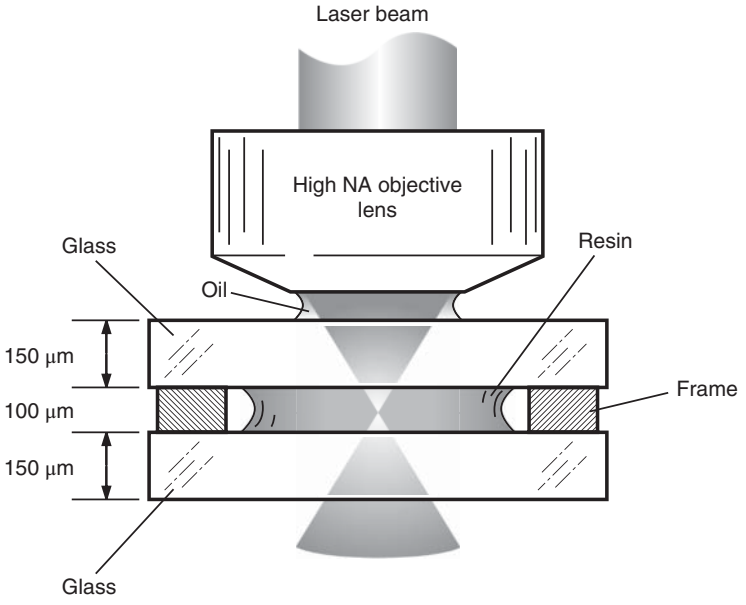


Figure 13.16 Setup for the two-photon polymerization process. Source: Adapted from Korte et al. (2004).

numerical aperture of 1.4 is used, resulting in a strongly convergent and thus tightly focused beam. The threshold time for solidification of the polymer may be about 0.2 ms. After the process, the unpolymerized resin can be removed by dissolving in acetone or ethanol, depending on the polymer.

The resolution of structures made by 2PP is determined by a number of factors:

1. Spherical aberration at the boundary between the sample and air due to their significantly different refractive indices. This causes deformation of the focused spot along the optical axis. This effect can be minimized by having oil between the lens and the sample to remove the refraction at the boundary between the sample and air.
2. The diffusion of radicals and polymer chain growth beyond the radiation zone. However, since 2PP is only effective in the focused regions of high intensity, the resin does not polymerize much beyond the focal volume where the photon density is high. Thus, the resolution is much better than would be obtained for the conventional single-photon absorption process.

13.2.2 Powder-Based Systems

There are different forms of the powder-based systems. Some have characteristics similar to the liquid-based systems described in the Section 13.2.1, while others are more similar to the solid-based systems to be discussed in Section 13.2.3. The common feature among these systems is that they use powder (which may be plastic or metallic) as the starting material. The major types of powder-based systems include the following:

1. SLS or powder bed fusion (PBF).
2. Direct metal deposition (DMD).
3. Binder jetting.

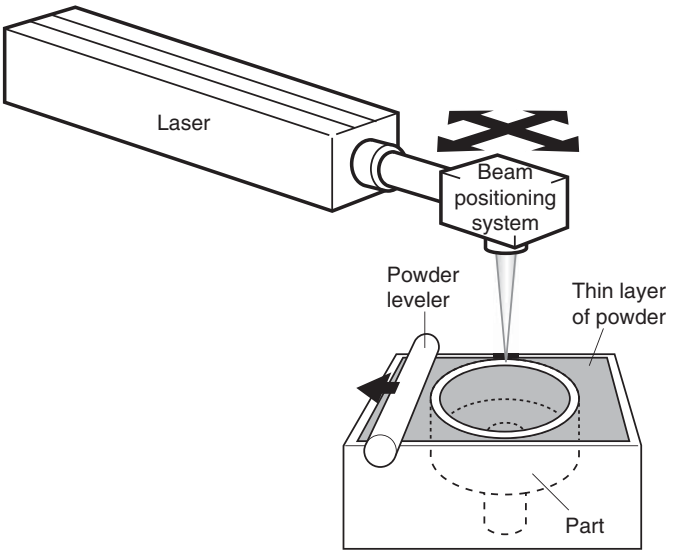


Figure 13.17 Schematic of the selective laser sintering process. Source: Kochan (1993)/with permission of Elsevier.

Only SLS and DMD are laser-based. Even though binder jetting does not involve the use of lasers, it is also discussed for the sake of completeness, and especially since it is becoming increasingly used.

13.2.2.1 Selective Laser Sintering (SLS) This process is now also widely known as the PBF method. It is similar in concept to the stereolithography approach and involves first depositing a thin layer of the powder material on the working platform within the process chamber. A high-power laser is then scanned on the powder layer to trace out the desired shape, heating up the scanned region to a temperature just below melting. The heat sinters the powder particles and thereby consolidates them, forming a solid mass. The beam is only directed to the regions defined by the part's designed shape, and the rest of the powder material remains unchanged. On completion of this layer, another layer of powder is deposited on top of it, using a roller mechanism (Figure 13.17). The process of applying a powder layer and sintering is repeated until the part is completed, after which the unbound powder is emptied from the rest of the part.

During the process, the powder that is not sintered serves as a built-in support structure. Thus, no separate supports need to be incorporated in the design, simplifying both the design and post-processing phases. However, the energy required may be about 300–500 times the energy required for photopolymerization of a layer of similar thickness. Also, an inert gas is used in the process chamber to prevent oxidation at the high temperatures. Various materials can be used, as long as they are in powder form, for example, thermoplastics, composites, ceramics, and metals. Sample conditions for additive manufacturing by SLS are shown in Table 13.3.

SLS has the advantage of being useful for a wide range of materials such as metals, ceramics, and plastics. Furthermore, different powders can be sintered at the same slice level, enabling a heterogeneous structure with specific desired characteristics to be fabricated. Any unsintered powder can be used again and is thus not wasted. The process is also one step and does not require any supports. However, the resulting accuracy is less than that of stereolithography, since it depends on the size of powder particles. Accuracy is further impaired by the tendency of neighboring particles to become attached to treated regions as a result of conduction. Also, internal hollow spaces that are created are

Table 13.3 Sample conditions for selective laser sintering.

Process parameter	Range of values	Units
Laser power	50–1000	W
Beam size	400	μm
Scan speed	0.9–2	m/s
Minimum layer thickness	20–500	μm

not as easily cleaned. Thus, components made by the SLS method may not be suitable for surgical applications. In addition, an inert atmosphere (usually argon) is needed to avoid oxidation during sintering. Another drawback of the process is that it is mainly effective for new parts built up on flat, 2D surfaces.

SLS machines are made by 3D Systems, EOS, and Sinterit, among others. They tend to be relatively more expensive than those for other AM processes and are thus mainly used in industrial applications. Earlier machines were based on the CO_2 laser. However, fiber lasers are now increasingly used because of their efficiency and higher absorption by metals.

13.2.2.2 Direct Metal Deposition (DMD) The same principle used for laser deposition (Section 11.3) can be used to create 3D objects by depositing one layer on top of another. One advantage of DMD is that it has the ability to control the type and proportions of material flow into the process, and thus the composition of the end product. Thus, functionally graded parts can be produced. It is also good for repairing damaged components. The major drawback is the relatively poor surface quality and resolution.

The two major forms of DMD currently in use are the following:

1. PBF, as discussed in Section 13.2.2.1.
2. Powder fed direct energy deposition (DED) method.

The procedure for the DED technique is essentially the same as the powder feeding method as discussed in Section 11.3.2. One major advantage is that it can be used with a five-axis arm and thus not limited to flat surfaces. It is good for repair work, surface modification, and hybrid manufacturing (combination of traditional metalworking processes and additive manufacturing). The two processes are compared in Table 13.4. Some of the companies that make DMD equipment are 3D Systems, EOS, Optomec, DM3D Technology, Renishaw, and Sinterit.

13.2.2.3 Binder Jetting The binder jetting process (originally known as 3D Printing) also starts with a layer of powder which is spread over a powder bed mounted on top of a piston and cylinder arrangement. However, this process is not laser-based, and in this case, the powder particles are consolidated by ejecting an adhesive bonding material onto the powder using technology similar to ink-jet printing, in the regions defined by the part's designed shape. The piston is then lowered, a new layer of powder is spread, and the binder applied (Figure 13.18). The process is repeated until the part is completed. The part is then subjected to heat treatment to further strengthen it, after which the unbound powder is removed. The powder may be metal, plastic, or ceramic. Binder jetting process has the advantage that the amount of material that is ejected is relatively small compared to the size of the build. Also, it is a lot easier to fabricate composite parts or heterogeneous components by adding color, for example. However, the process is slowed by the fact that the powder needs to be distributed.

Binder jetting machines are produced by ExOne, Voxeljet, Microjet Technology, and 3D Systems.

Table 13.4 Comparison of the powder bed fusion and direct energy deposition methods.

Powder bed fusion (PBF)	Directed energy deposition (DED)
Smaller build volume – used for small or nested parts	Larger build volume – no restriction on part size
Slower build speed (0.06 kg/h for steel)	Faster build speed (0.5 kg/h for steel with LENS 2 kW laser)
Used to build new parts – limited ability to add material to an existing part, and requires flat substrate	Used to build new parts, repair parts, and add wear-resistant surfaces on flat or 3D substrates
Limited multiple material printing capability	Programmatically blend or grade multiple materials
Complex geometry with high resolution (Ra: 9–12 μm)	Simple geometry with less resolution (Ra: 20–50 μm)
Restricted powder availability/higher powder cost	Commercially available powders/lower powder cost

Source: From Vartanian and Pierra (2018)/Endeavor Business Media.

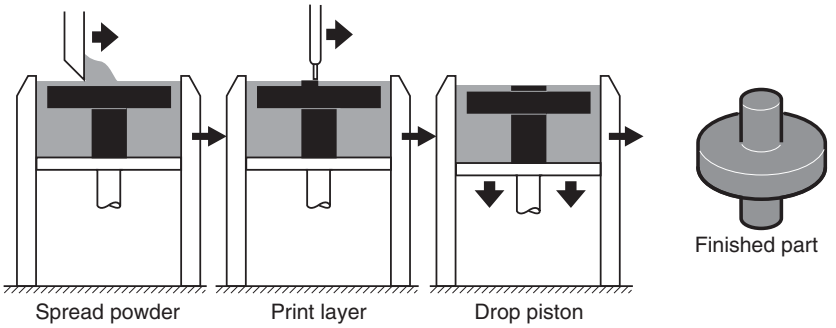


Figure 13.18 Principle of the binder jetting process. Source: Kochan (1993)/with permission of Elsevier.

13.2.3 Solid-Based Systems

There are several forms of the solid-based systems, each differing in the manner in which the layer is applied. However, due to space constraints, we shall only discuss three of the most common systems. These are as follows:

1. Fused deposition modeling (FDM).
2. Laminated object manufacturing (LOM).
3. Wire deposition

LOM is laser-based, but FDM is not. Others are the selective adhesive and hot press system, the multi-jet modeling system, the model maker MM-6B, and the hot plot.

13.2.3.1 Fused Deposition Modeling With this process, which is not laser-based, the material is supplied as a filament wire (about 1.27 mm in diameter) on a spool. It is fed through a heated FDM head, where it becomes softened into a viscous state and then is extruded and deposited in ultra-thin layers (Figure 13.19). The extruded material quickly solidifies after deposition. The horizontal width of the extruded material can vary between 0.25 and 2.50 mm, and the layer height between 50 and

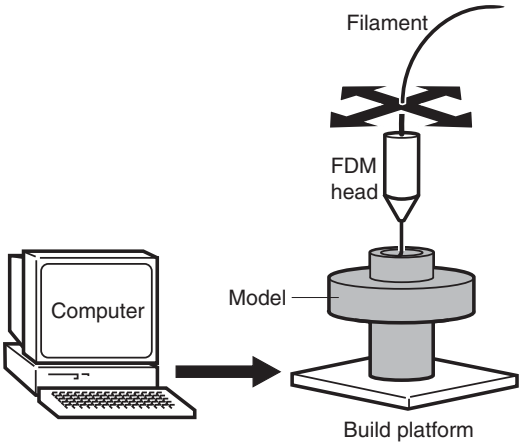


Figure 13.19 Principle of the fused deposition modeling process. Source: Kochan (1993)/with permission of Elsevier.

400 μm. As with the other additive manufacturing processes, when one layer is completed, the next layer is laid on top of it until the part is done. Thermoplastics, including aerospace grades such as polyetherimide (PEI) and polyetheretherketone (PEEK), are typical materials that are processed by this method. Materials that are used for everyday consumer items include polylactide (PLA) and acrylonitrile butadiene styrene (ABS). However, FDM machines have lower resolution and production rates compared to SL systems. FDM machines are more widely used than SLS systems, since the SLS equipment tend to be more expensive and are thus mainly used in industry. FDM systems are produced by several companies, including Stratasys and HP. Some of the desktop products currently on the market include LulzBot, RepRap, and Ultimaker. The typical build size of a desktop printer is about 0.2 m × 0.2 m × 0.2 m, and that of an industrial machine may be about 1 m × 1 m × 1 m.

FDM systems that are capable of producing composite-based products are being developed by Stratasys, Impossible Objects, and Markforged.

13.2.3.2 Laminated Object Manufacturing The principle of this process is illustrated in Figure 13.20 and involves sequentially bonding together thin layers of material. There are different ways that bonding can be accomplished:

1. Using an adhesive-coated sheet material that is pressed to the preceding sheets.
2. Using ultrasonic welding to bond the current sheet to the previous layers.
3. Thermal bonding.
4. Mechanical clamping.

With the adhesive-coated method, the sheet material is obtained from a supply roll. Once the sheet is in place, a laser beam is directed to cut the outline of the part in that sheet. Cross-hatches are then made by the beam in the excess material outside the part boundary, turning it into a support structure with a rectangular boundary that remains in place. After that, the platform is lowered with the formed stack, and a fresh section of the sheet material is moved on top of it. The platform is then raised, and a heated roller rolls over it, bonding it to the preceding layer in one reciprocal motion. The part outline is then cut with the laser beam. This continues until the entire part is produced. The entire rectangular block then comes out, and the part is separated from the surrounding material

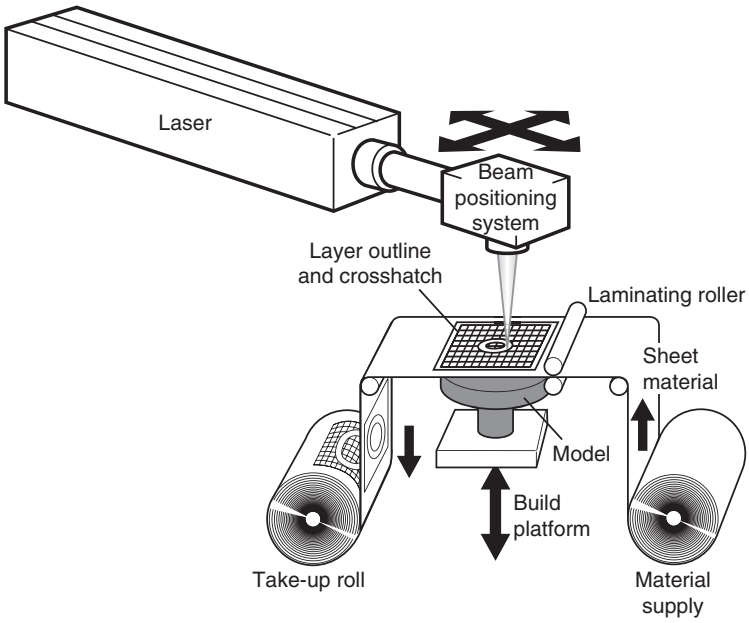


Figure 13.20 Principle of the laminated object manufacturing process. Source: From Chua and Leong (1997).

Table 13.5 Sample conditions for laminated object manufacturing.

Process parameter	Range of values	Units
Laser power	25–50	W
Beam size	250–400	μm
Scan speed	0.3–0.7	m/s
Material thickness	0.05–0.4	mm

Source: From Chua and Leong (1997).

which is easily removed as a result of the cross-hatching. Sample conditions for laminated object manufacturing are shown in Table 13.5.

When using ultrasonic bonding, a disk-shaped sonotrode rolls along the sheet metal under pressure, while oscillating at ultrasonic frequencies (typically 20 kHz) in a direction parallel to the cylindrical axis with an amplitude of about 40 μm. Each layer consists of a series of foils which are laid next to each other and bonded to the preceding layer one after the other. A milling machine is often used to trim the part after about four layers have been bonded. Foil thicknesses are typically about 150 μm. The production of complex structures, for example honeycombs and cooling channels, are facilitated using this technique. Also, it is a lot easier to embed fibers in the structure.

Thermal bonding uses heat to consolidate one layer with the next, while mechanical coupling involves the use of mechanical fastening devices.

Laminated object manufacturing is useful for a wide range of materials, including paper, plastics, metals, ceramics, and composites. However, it has the disadvantage that unused material resulting from internal hollow spaces can only be removed with difficulty. Also, finishing costs are generally higher than for the other laser-based processes.

An alternate approach to laminated object manufacturing combines both additive and subtractive processes where relatively thick slices are machined to shape and bonded to each other to form the

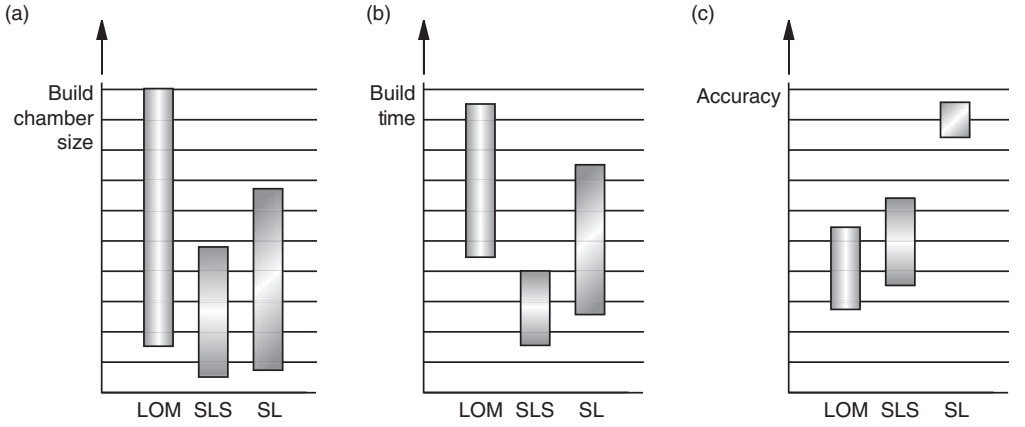


Figure 13.21 Qualitative comparison of laser-based additive manufacturing systems (a) build chamber size (b) build time (c) accuracy. Source: From Gebhart (2003).

3D object. This approach is useful for very large products that may be difficult to manufacture using conventional machining. The bond strength determines the application for which the product will be suitable. Bonding methods range from gluing to diffusion welding in metals.

The laminated object manufacturing method has also been used to produce laser-induced graphene foams. This involves irradiating commercial polyimide film with a CO_2 laser to convert the polyimide into a layer of porous graphene film. One layer is used as the base of a graphene foam, while the other layers are coated with ethylene glycol as an adhesive, stacked above each other, and fused together using the laser. The printed foam is then heated at 200 °C and atmospheric pressure or in vacuum at 600 °C to evaporate any remaining ethylene glycol.

Overall, LOM is not as commonly used as some of the other additive manufacturing methods. Variations of machines based on the LOM concept are produced by Cubic Technologies and Mcor Technologies.

13.2.3.3 Wire Deposition The wire deposition technique is essentially an enhancement of the gas metal arc welding (GMAW)/metal inert gas (MIG) process. Variations of the gas tungsten arc, plasma arc, and shielded metal arc welding processes can also be used. It basically involves deposition of weld beads, layer by layer, to generate a desired design.

13.2.4 Qualitative Comparison of Some Major Systems

In this section, some of the major additive manufacturing techniques are compared qualitatively (Figure 13.21). Quantitative comparison is difficult to achieve since a number of factors (such as system calibration, operator expertise, and method of measurement) do influence the outcome.

13.3 POST-PROCESSING

Depending on the build methodology, post-processing may involve such operations as removal of the part, cleaning, post-curing, and part finishing. The major post-processing operations and the processes for which they are normally needed are summarized in Table 13.6.

For the liquid-based systems, once the part is completed, excess resin is allowed to drain back into the vat, and any resin adhering to the part wiped off. A solvent is then used to clean it and low-pressure

Table 13.6 Post-processing operations.

	Stereolithography	Selective laser sintering	Laminated object manufacturing
Post-curing	•		
Cleaning	•	•	•
Part finishing	•	•	•

compressed air subsequently used to dry it. Finally, the part is post-cured using broadband ultraviolet radiation to complete the polymerization process and increase the strength of the part. This last step may take a few hours depending on the size of the part.

Finishing is often necessary because of factors that affect the process such as shrinkage, distortion, and curling and depends on the intended application. Common finishing operations to which the part may be subjected include cutting, sandblasting, polishing, furnace sintering, coating, and painting.

1. *Cutting*: Traditional cutting processes such as turning, grinding, milling, and drilling may be used to improve the surface finish of the part, remove unwanted supports, introduce additional form features, or improve dimensional accuracy.
2. *Sandblasting and Polishing*: Sandblasting may also be used to clean up the part, but for better appearance and higher accuracy, polishing would be necessary.
3. *Coating and Painting*: Coating is often used to provide a protective layer for the part, while painting is normally used for esthetic purposes.

13.4 APPLICATIONS

Early additive manufacturing (AM) systems were developed primarily for "touch-and-feel" applications that provided designers a means to visualize how the final product would look, without regard to their function. In recent years, as newer techniques have been developed, the range of applications has increased considerably. The major areas are as follows:

1. Design.
2. Engineering, analysis, and planning.
3. Manufacturing and tooling.
4. Personalized production.

13.4.1 Design

Early applications of additive manufacturing in design included such issues as verification of CAD models, object visualization, proof of concept, and marketing. Additive manufacturing enables designers to have a physical part that confirms the design that they have created using CAD. This is especially important when esthetic requirements are critical. A number of engineering components are quite complex, physically. Attempting to visualize the final form using CAD drawings can be quite challenging even for experienced designers, and more so for other engineers and management who may need to be involved in final design approval. The availability of a prototype significantly facilitates discussion among these groups. Finally, for marketing purposes, a physical model is

always beneficial, enabling potential customers to get a feel for the final product. The prototype can also be used for marketing brochures.

13.4.2 Engineering, Analysis, and Planning

Aspects of engineering, analysis, and planning for which additive manufacturing is useful include scaling, form and fit, stress and flow analyses, mock-up parts, planning surgical operations, and custom prostheses and implants. AM techniques enable models to be scaled up or down by scaling the CAD design. When the appropriate size is selected, the CAD data is changed accordingly for creating the AM model. How a part or component fits in with the overall design is something that can be assessed using a prototype.

The flow characteristics associated with products that involve fluid flow are more easily evaluated with the aid of a prototype. Likewise, prototypes facilitate the evaluation of stresses that can develop in a part, using photo-optical methods. Furthermore, prototypes or mock-up parts can be assembled into the complete product and used for final testing. Finally, AM technology can also be used to produce custom prostheses and implants, as well as for planning surgical and reconstructive procedures.

13.4.3 Manufacturing and Tooling

In the early years, most applications of AM in manufacturing and tooling involved making patterns to be used in various processes, for example metal casting, metal spraying, and plastic vacuum casting with silicon molding. We shall illustrate the use of AM in these areas with metal casting.

Metal casting involves making a mold which has a cavity into which molten metal is poured and allowed to solidify. The mold is then opened and the casting is taken out. The mold may be made of sand, plaster, or metal. For making a sand or plaster mold cavity, a pattern is often required. The pattern is made to the shape of the part to be produced, with only slight modifications, and is usually made out of wood, metal, or plastic material. Metal and plastic patterns can be made using additive manufacturing. Metal molds can also be made directly using additive manufacturing. We shall illustrate the application of AM in metal casting with the investment casting (lost wax) process (Figure 13.22).

Traditionally, this process first involves preparing a pattern by injecting wax or thermoplastic resin into a die. A number of separate subpatterns may be joined together after they are removed from their respective dies, using an adhesive, to form a more complex pattern. A flask is then placed around the pattern, and ceramic slurry poured into the flask to cover the pattern, to form the mold. After the mold has set, it is heated to melt out the wax, leaving the mold cavity, into which the molten metal can be poured for casting. Additive manufacturing can be used to make the pattern faster and more economically.

13.4.4 Personalized Production

In recent years, rapid prototyping has evolved into additive manufacturing (or 3D Printing), enabling customized end products to be made. The cost of the production systems has reduced to the point where individuals can now purchase their own 3D Printing equipment for simple custom products. The technology has also evolved to the point where products can be made at a reasonable rate. Such applications include medical (3D bioprinting), aerospace, automotive, architectural, and medical industries. However, it is still an open question if additive manufacturing can be used for mass production.

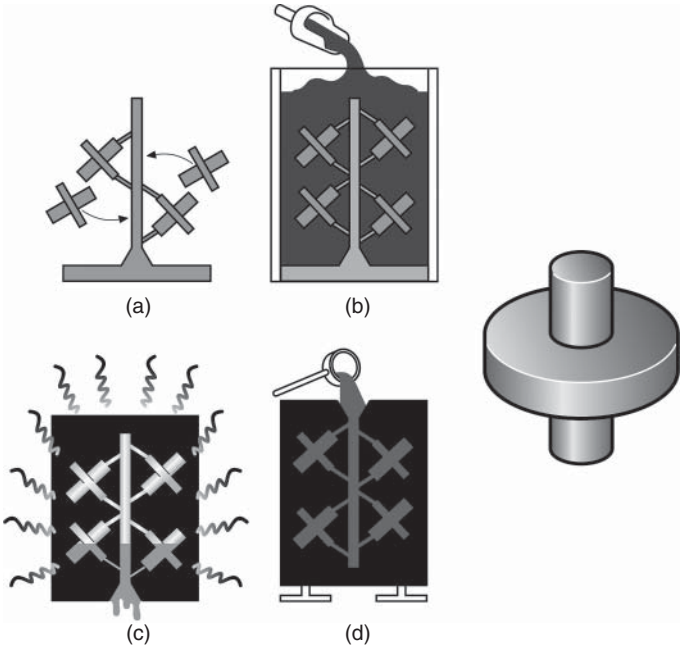


Figure 13.22 The investment casting process. (a) Making basic unit of pattern. (b) Joining pattern units into a more complex pattern. (c) Placing the flask. (d) Filling flask with mold material. (e) Melting off the pattern. (f) Pouring the molten material into the mold. Source: From Lindberg (1977).

13.5 ADVANTAGES AND DISADVANTAGES

13.5.1 Advantages

1. It eliminates the need to consider the production method while designing the product, thereby making the design process simpler.
2. It helps reduce the overall time required for product development.
3. Production of the part is mostly done with one process as opposed to the multiple processes required when using conventional manufacturing methods.
4. It makes it possible to customize products to the end user needs.

13.5.2 Disadvantages

1. The production rates achievable with additive manufacturing are such that the process is currently only good for low volume production.
2. The surface quality achievable with the process is low compared to traditional manufacturing methods.

13.6 SUMMARY

Additive manufacturing involves the generation of physical objects from graphical computer data and is used in design conceptualization, to facilitate engineering stress analysis, and in making

patterns and/or dies for manufacturing operations. Making a part by additive manufacturing involves three stages:

1. *Computer-Aided Design*: A CAD package is used to develop a computer model of the part to be produced. Where a physical object already exists, reverse engineering may be used to develop a computer model from the object. The CAD model is next converted into a format required by the additive manufacturing (AM) equipment. A commonly used format is the STL format. For processing, the CAD model is sliced into cross sections to facilitate the layer-by-layer building of the part.
2. *Part Building* (*Processing*): Different materials can be processed by additive manufacturing. The process can thus be categorized by the type of material, viz. liquid-based, powder-based, or solid-based system. In each case, the part is built one layer at a time. The liquid-based systems generally involve the use of lasers. One powder-based system that uses lasers is the SLS process. One solid-based system, laminated object manufacturing, also uses a laser.
3. *Post-Processing*: After the part is made, it may be further finished by machining, sand-blasting and polishing, and/or coating and painting.

Finally, 3D micro-components that have nano-features can also be created using 2PP by tightly focusing a femtosecond infrared laser into a photosensitive resin.

APPENDIX 13.A

List of symbols used in the chapter

Symbol	Parameter	Units
a_i, a_{ij}	Polynomial coefficients	—
$B_{n,i}$	Bezier polynomial	—
C_d	Cure depth	m
D_p	Penetration depth	m
n	Degree of polynomial	—
P_i	Points to be fitted with a polynomial	—
Q_c	Critical exposure or energy per unit area	J/m^2
Q_0	Incident energy per unit area	J/m^2
S_j	Span j of spline	—
t	Dummy variable	—
W_m	Maximum width of cured region	m
x_i	Data points	—
y_c, z_c	Values of y, z when $Q_\mathrm{p} = Q_\mathrm{c}$	m

PROBLEMS

13.1. Determine the nonparametric natural cubic spline that fits the following data points:
(0, 0), (2, 3), (3.5, 2)

13.2. Obtain the Bezier curve that passes through the following data points:
(0, 2), (2, 3), (3.5, 6), (5, 1)

13.3. Obtain a natural cubic spline that fits the following data:
$P_1[0, 5]$, $P_2[1, 2]$, $P_3[2, -3]$, $P_4[3, -4]$, and $P_5[4, 5]$

13.4. For a Gaussian beam, the intensity of the incident beam, I_0, on the surface of the liquid layer is given by Eq. (13.15) as:

$$I_0(r) = I_\mathrm{p} e^{\left(-\frac{2r^2}{w^2}\right)}$$

show that the peak intensity, I_p, is given by:

$$I_\mathrm{p} = \frac{2q}{\pi w^2}$$

13.5. Given the surface intensity of a laser beam on a slice:

$$I_0(r) = \frac{2q}{\pi w^2} e^{\left(-\frac{2r^2}{w^2}\right)}$$

show that the corresponding surface exposure, Q_0, becomes

$$Q_0(y) = \sqrt{\frac{2}{\pi}} \frac{q}{w u_x} e^{\left(-\frac{2y^2}{w^2}\right)}$$

13.6. The cross section of a scanned line is given by:

$$Q_\mathrm{c} = \sqrt{\frac{2}{\pi}} \frac{q}{w u_x} e^{\left(-\frac{2{y_\mathrm{c}}^2}{w^2}\right)} e^{-\frac{z_\mathrm{c}}{D_\mathrm{p}}}$$

show that the above equation reduces to the form:

$$z_\mathrm{c} = a{y_\mathrm{c}}^2 + b$$

13.7. Given that the parabolic profile of the line-scan cross section is

$$z_\mathrm{c} = a{y_\mathrm{c}}^2 + b$$

show that the maximum cure depth that results from a single line scan is

$$C_\mathrm{d} = b = D_\mathrm{p} \ln \left[\sqrt{\frac{2}{\pi}} \left(\frac{q}{w u_x Q_\mathrm{c}} \right) \right]$$

13.8. Given that the parabolic profile of the line-scan cross section is

$$z_\mathrm{c} = a{y_\mathrm{c}}^2 + b$$

show that the maximum cured width (on the surface of the slice) that results from a single line scan is

$$W_\mathrm{m} = w \sqrt{2 \frac{C_\mathrm{d}}{D_\mathrm{p}}}$$

13.9. An argon ion laser of output power 250 mW is used to process a photo-curable resin at a scan speed of 2.5 m/s, with a beam radius of 250 μm. If the resin has a critical exposure of 13.5 mJ/cm^2 and the depth, z_c, at a distance of $y_c = 0.1$ mm from the scan line on the surface is 0.32 mm, determine the cure depth, C_d, for the process.

13.10. For the conditions given in Problem 13.9, determine

The maximum cured linewidth.

The equation of the cross section for the cured region.

The peak intensity, I_p, of the incident beam.

13.11. The object shown in the figure below is to be made using stereolithography.

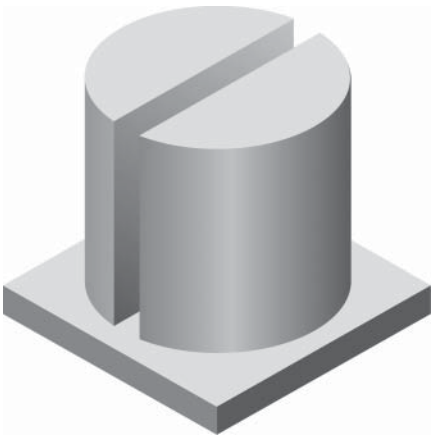


Figure P13.11

(a) Sketch the tree structure with the Boolean operations (union, intersection, and difference) necessary for generating a CAD of the object.

(b) In building the object, the photocurable resin is scanned with a 500 mW laser beam at a speed of 1 m/s. The beam radius is 150 μm. The critical exposure of the resin is 15 mJ/cm^2. If $D_p = 1$ mm, determine the depth from the scan line at a distance of $y_c = 0.15$ mm. Calculate the corresponding cure depth C_d and the maximum width of the cured region W_m.

BIBLIOGRAPHY

Amirouche, F.M.L. (1993). *Computer-Aided Design and Manufacturing*. Englewood Cliffs, NJ: Prentice-Hall.

Belfield, K.D. (2001). Two-photon organic photochemistry. *The Spectrum* 14 (2): 1–7.

Billmeyer, F.W. Jr. (1962). Radical chain (addition) polymerization. In: *Textbook of Polymer Science*, 49–81. New York: Wiley Interscience.

Chua, C.K. and Leong, K.F. (1997). *Rapid Prototyping: Principles & Applications in Manufacturing*. New York: Wiley.

Cooper, K.G. (2001). *Rapid Prototyping Technology*. New York: Marcel Dekker, Inc.

Davies, B.L., Robotham, A.J., and Yarwood, A. (1991). *Computer-Aided Drawing and Design*. London: Chapman & Hall.

Galajda, P. and Ormos, P. (2001). Complex micromachines produced and driven by light. *Applied Physics Letters* 78 (2): 249–251.

Gebhart, A. (2003). *Rapid Prototyping*. Munich: Hanser Publishers.

Gibson, I., Rosen, D., and Stucker, B. (2015). *Additive Manufacturing Technologies - 3D Printing, Rapid Prototyping, and Direct Digital Manufacturing*, 2e. New York, NY: Springer.

Gregory, S.L.L.H.F. (2012). Analysis of two-photon polymerization for the development of a reconfigurable micro-device manufacturing system. PhD Thesis. University of Michigan.

Gu, D.D., Meiners, W., Wissenbach, K., and Poprawe, R. (2012). Laser additive manufacturing of metallic components: materials, processes and mechanisms. *International Materials Reviews* 57 (3): 133–164.

Jacobs, P.F. (1992). *Rapid Prototyping & Manufacturing: Fundamentals of StereoLithography*. Dearborn, Michigan: Society of Manufacturing Engineers.

Kawata, S., Sun, H.-B., Tanaka, T., and Takada, K. (2001). Finer features for functional microdevices. *Nature* 412: 697–698.

Kochan, D. (1993). *Solid Freeform Manufacturing*. Amsterdam, Holland: Elsevier.

Korte, F., Koch, J., Serbin, J. et al. (2004). *IEEE Transactions on Nanotechnology* 3 (4): 468–472.

Kruth, J.P., Leu, M.C., and Nakagawa, T. (1998). Progress in additive manufacturing and rapid prototyping. *Annals of the ClRP* 47 (2): 525–540.

Levy, G.N., Schindel, R., and Kruth, J.P. (2003). Rapid manufacturing and rapid tooling with layer manufacturing (LM) technologies, state of the art and future perspectives. *CIRP Annals Manufacturing Technology* 52 (2): 589–609.

Lindberg, R.A. (1977). *Processes and Materials of Manufacture*. Boston, MA: Allyn and Bacon, Inc.

Luong, D.X., Subramanian, A.K., Silva, G.A.L. et al. (2018). Laminated object manufacturing of 3D-printed laser-induced graphene foams. *Advanced Materials* 30: https://doi.org/10.1002/adma.201707416.

Maruo, S. and Kawata, S. (1998). Two-photon-absorbed near-infrared photopolymerization for three-dimensional microfabrication. *Journal of Microelectromechanical Systems* 7 (4): 411–415.

Maruo, S., Nakamura, O., and Kawata, S. (1997). Three-dimensional microfabrication with two-photon-absorbed photopolymerization. *Optics Letters* 22 (2): 132–134.

Onwubiko, C. (1989). *Computer-Aided Design*. St. Paul, MN: West Publishing Company.

Pham, D.T. and Dimov, S.S. (2001). *Rapid Manufacturing*. London: Springer-Verlag.

Robinson, K. (2002). Two-photon absorption enables microfabrication. *Photonics TechnoWorld*, May 2002 Edition.

Strickler, J.H. and Webb, W.W. (1990). Two-photon excitation in laser scanning fluorescence microscopy. *SPIE CAN-AM Eastern* 1398: 107–118.

Sun, H.-B., Matsuo, S., and Misawa, H. (1999). Three-dimensional photonic crystal structures achieved with two-photon-absorption photopolymerization of resin. *Applied Physics Letters* 74 (6): 786–788.

Sun, H.-B., Kawakami, T., Xu, Y. et al. (2000). Real three-dimensional microstructures fabricated by photopolymerization of resins through two-photon-absorption. *Optics Letters* 25 (15): 1110–1112.

Sun, H.-B., Takada, K., and Kawata, S. (2001a). Elastic force analysis of functional polymer submicron oscillators. *Applied Physics Letters* 79 (19): 3173–3175.

Sun, H.-B., Tanaka, T., Takada, K., and Kawata, S. (2001b). Two-photon photopolymerization and diagnosis of three-dimensional microstructures containing fluorescent dyes. *Applied Physics Letters* 79 (10): 1411–1413.

Sun, H.-B., Tanaka, T., and Kawata, S. (2002). Three-dimensional focal spots related to two photon excitation. *Applied Physics Letters* 80 (20): 3673–3675.

Vartanian, K. and Pierra, P. (2018). Is welding the Trojan horse of metal additive manufacturing? *Industrial Laser Solutions* 27–29.

Wong, K.V. and Hernandez, A. (2012). A review of additive manufacturing. *International Scholarly Research Network, ISRN Mechanical Engineering* 2012: 208760, 10 pages, https://doi.org/10.5402/2012/208760.

14 Medical and Nanotechnology Applications of Lasers

Lasers have found extensive applications in the medical field over the years. With the field of nanotechnology being relatively recent, new developments are constantly evolving. Even though some of the nano-applications are in the medical field, we separate the discussion into two sections. In the first section, we focus on medical applications of lasers, limiting the discussion to medical device manufacturing and therapeutic applications.

Nanotechnology applications are presented in the second section. We start by discussing the various ways in which nano-features or structures can be produced. This is then followed by a discussion on a number of the different features that can be produced, specifically nanoholes and gratings, followed by nanobumps. Finally, laser-assisted nanoimprint technology is presented.

14.1 MEDICAL APPLICATIONS

Laser applications in the medical field cover a number of areas. These may be broadly categorized as:

1. Medical device manufacturing.
2. Therapy.
3. Diagnostics.

There are several applications in each of these areas. However, due to space limitations, we shall only discuss the manufacturing and therapeutic applications, highlighting only one or two examples in each of those areas.

The use of the laser as a therapeutic tool and as a diagnostic tool in medicine is summarized in Tables 14.1 and 14.2, respectively.

In the next section, we shall discuss the application of lasers in the fabrication of medical devices, using stent manufacturing to illustrate the application.

14.1.1 Medical Devices

A number of medical devices are fabricated using lasers. For metallic components, the basic principles discussed in preceding chapters do apply. We shall illustrate these with one specific example, that is, stents (Figure 14.1a). A stent is a small, lattice-shaped, metal tube that is inserted permanently into an artery. It provides a minimally invasive method of treating coronary artery diseases such as heart attacks and strokes. These diseases are normally a direct result of limited blood supply due to

Principles of Laser Materials Processing: Developments and Applications, Second Edition. Elijah Kannatey-Asibu, Jr.
© 2023 John Wiley & Sons, Inc. Published 2023 by John Wiley & Sons, Inc.
Companion Website: www.wiley.com/go/Kannatey-Asibu/PrinicplesLaserMaterialsProcessing

Table 14.1 Therapeutic applications of the laser.

	Disruption	Coagulation	Cutting	Ablation
Surgery		•	•	
Gynaecology		•	•	
Urology	•	•		
Ortholaringology		•	•	•
Opthalmology	•	•		•
Dentistry		•	•	•
Orthopedics		•		•
Gastroscopy		•	•	
Dermatology	•	•		•

Table 14.2 Diagnostic applications of the laser.

	Fluorescent spectroscopy	Doppler spectroscopy	Optical tomography
Tumor recognition	•		•
Blood throughput		•	
Tissue differentiation	•		•
Tissue structures			•
Metabolic activity	•		

the constriction of blood vessels. Gradual buildup of fats in the arteries is believed to be the main cause for arterial blockage. Earlier treatment of such blockage involved bypass surgery. The advent of stents in the 1980s began with balloon angioplasty and provided a less expensive and less traumatic alternative to bypass surgery for some patients. This evolved into the use of metallic stents, which are more durable. The stent helps to open an artery so that blood can flow through it. The treatment simply involves placing the stent onto a balloon catheter assembly and inserting into the artery. The balloon is then inflated (Figure 14.1b) forcing the artery to open up. When the balloon is subsequently deflated and removed, the stent is left in place, keeping the artery open.

Such a procedure is normally good for about six months, after which growth of muscle cells at stent–artery wall interface may lead to arterial reblockage, a phenomenon referred to as restenosis. Among other things, restenosis has been linked to the smoothness of finished stents, as investigations have shown that cells that cause restenosis preferentially grow at finished surface and arterial wall interfaces. In recent years, drug-eluting stents have been developed to minimize the onset of restenosis. These are very effective. However, they are also relatively expensive, compared to bare-metal stents. Furthermore, recent studies have shown that drug-eluting stents may result in higher incidence of stroke or heart attack compared to bare-metal stents.

Manufacturing of metallic stents involves the following steps:

1. Drawing of a tubing with dimensions that match the final stent diameter and wall thickness. Typical stent diameters range from 2 to 5 mm, with wall thicknesses ranging from 50 to 200 μm, and lengths from 6 to 38 mm.
2. *Laser Cutting*: the appropriate design, as illustrated in Figure 14.1c, is then cut out of the tubing, usually with a laser. Traditional laser cutting results in the formation of striations and dross. Typical surface roughness obtainable with conventional laser cutting is in the range 0.8–6.3 μm. Finishing operations are therefore necessary to ensure a desirable surface quality.

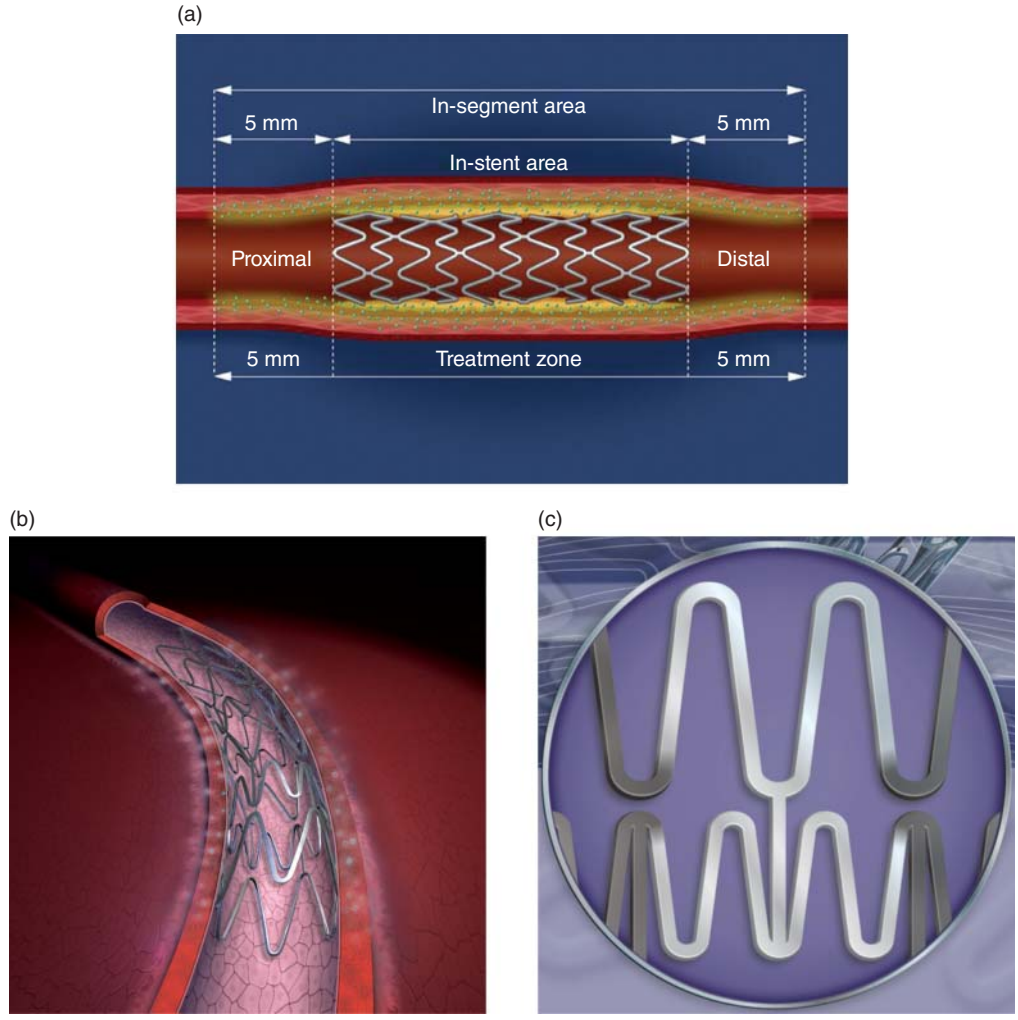


Figure 14.1 Coronary stent. (a) Multi-link stent on a balloon. (b) Uniform scaffolding multi-link stent. (c) General multi-link stent. Source: By permission of Boston Scientific Corporation.

3. Finishing operations as commonly performed, in order, are
   a. Pickling, which involves chemical etching done in acid solution to remove most of the dross.
   b. Electropolishing, to remove any residual dross or sharp points that may be left. This involves passing electric current through an electrochemical solution.
4. The stent is finally placed onto a balloon catheter assembly, sterilized, and packaged.

Materials that have traditionally been used for stents include stainless steels (typically 316L series), cobalt-based alloys, tantalum alloys, nitinol (nickel and titanium) shape memory alloy, and biocompatible materials such as polyhydroxybutyrate (PHB).

A number of lasers such as CO_2, Nd:YAG, excimer, diode, copper vapor, and titanium sapphire lasers have been used for cutting 316L-series stainless steel material in stent manufacturing. With the

traditional continuous wave or long pulse lasers, the thermal energy associated with the laser beam affects the microstructure of the material in the heat-affected zone (HAZ), and thereby the mechanical properties of the stent. This is undesirable for the following reasons:

1. Stents are permanently deformed by stretching during deployment in angioplasty.
2. Radial pressure from arterial walls and pressure from heart beats subject the stent to fatigue.

To mitigate these problems, ultrashort pulse lasers (pulse duration in the range 10^{-15} to 10^{-11} seconds) are being investigated for stent fabrication. These are capable of significantly improving the surface quality and thereby eliminating the need for additional finishing operations. Also, due to the reduced heat input and the short time available for the laser beam, the HAZ size is minimized.

14.1.2 Therapeutic Applications

There are several therapeutic applications of the laser, and due to space limitations, we shall only discuss a few of these, specifically in surgery, opthalmology, dermatology, and dentistry. In each of these areas, the procedure may be classified under photocoagulation, photodynamic therapy, photodisruption, photoevaporation, or photoablation. In a number of these applications, lasers have several advantages over traditional procedures. These include

1. Hemostasis or the ability to minimize hemorrhaging through coagulation and thereby provide a clear surgical field.
2. Precision of the process.
3. Minimal damage to collateral tissue.
4. Rapid and relatively painless healing.
5. The ability to control the beam intensity, enabling control of depth and extent of procedure.
6. The ease with which the laser beam can be positioned and also used to access difficult to reach areas.

14.1.2.1 Surgical Procedures As Table 14.1 indicates, laser-based surgical procedures primarily involve coagulation and material removal, where they have significant advantages over mechanical incision.

One common surgical procedure that uses a laser is laser-assisted uvulopalatoplasty (LAUP). This is used to treat snoring which results from palatal flutter by reducing the tissues of velum and uvula and stiffening them. This minimizes or eliminates obstruction of airway, as well as vibration of soft tissue at the soft palate level. It has the advantage of fewer long-term complications compared with the traditional uvulopalatopharyngoplasty. The procedure involves using the laser as either an incision tool to cut material or as a heat source to vaporize material. It is normally performed using local anesthesia.

Most laser systems can be used for this procedure. The CO_2 laser is quite commonly used, even though diode and Nd:YAG lasers are also extensively used. The latter two have the advantage of fiber transmission, with the diode laser having the added advantage of compactness. However, both the Nd:YAG and diode lasers tend to result in a deeper coagulation zone and usually take longer to heal, compared to the CO_2 laser. The postsurgical pain associated with the process is also more severe for the Nd:YAG and diode lasers.

14.1.2.2 Opthalmology Lasers are ubiquitous in opthalmology and constitute a significant portion of opthalmic therapy. Our focus here will be on photocoagulation and photodynamic applications.

As the name implies, photocoagulation involves the use of a laser (light source) to coagulate (clot) or destroy abnormal, leaking blood vessels thereby stopping their growth. The process relies on the selective absorption of light around the green wavelength range by hemoglobin, the pigment in red blood cells. Various lasers are used for this procedure in opthalmology, for example, the argon ion, krypton ion, diode, Nd:YAG, and dye lasers. They are usually used in the CW mode with maximum output power of 2 W, but only for short periods of about 0.1–1.0 seconds and with focal diameters of about 0.05–2 mm.

One common application of the procedure is in diabetic retinopathy where it is used to seal leaking blood vessels in the retina, slowing their growth and thereby reducing the risk of vision loss. It is an outpatient procedure that uses local or topical anesthesia. Other applications of photocoagulation include treatment of detached retina, tumors of the retina, and age-related macular degeneration, which is a leading cause of blindness in the elderly.

In photodynamic therapy, a photosensitizer dye, which is a type of drug, is first injected into the patient. The dye has the tendency to accumulate in tumor tissue. When the tissue is now irradiated with light that is selectively absorbed by the photosensitizer, it produces a type of oxygen that induces cytotoxicity in the tissue cells, destroying them. It is usually an outpatient procedure and also has application in the treatment of age-related macular degeneration. Photodynamic therapy is also known by other names as phototherapy, photoradiation therapy, and photochemotherapy.

14.1.2.3 Dermatology Perhaps the most widely known use of lasers in medicine is in dermatology where they are used in skin resurfacing to improve skin appearance, in the treatment of skin tumors, in the treatment of scars and keloids, in the removal of tattoos, and so on.

The lasers that are most commonly used in laser skin resurfacing are the CO_2 and Er:YAG lasers. The coefficient of absorption in water of the Er:YAG laser beam is much higher than that of the CO_2 laser. Thus, for the Er:YAG laser, a significant amount of laser energy is absorbed in a thin layer of about 30 μm with a thermal damage zone of about 50 μm, compared to about a 100 μm depth and 150 μm damage zone for the CO_2 laser. Thus, there is less risk of skin damage with the Er:YAG laser. However, since it is necessary to penetrate to depths of about 250–400 μm for effective laser skin resurfacing, about three passes are required for the CO_2 laser as opposed to the 12–15 passes required for the Er:YAG laser. Also, the outcome for deep wrinkles is better with the CO_2 laser than the Er:YAG laser. For facial resurfacing, general or local anesthesia may be used depending on the extent of the procedure.

A number of skin malignancies or tumors are treated using a variety of phototherapies, including photodynamic therapy, photovaporization, and photocoagulation. In photodynamic therapy, tunable dye lasers are commonly used, even though other forms of noncoherent light are becoming increasingly popular. CO_2 lasers are more extensively used in procedures requiring photovaporization. The beam may be defocused or focused, with the defocused beam being used to provide homogenous vaporization of superficial layers, while the focused beam is mainly used for tissue incision. A CW beam is normally used for large skin lesions, while a superpulse with a pulse duration ranging between 50 and 200 ms is preferred for small lesions. Power levels typcially range from 5 to 20 W, with a beam diameter of about 2 mm. Photocoagulation for skin therapy is usually done with an Nd:YAG laser.

14.1.2.4 Dentistry The use of lasers in dentistry has largely been limited to soft tissue applications. Soft tissues include the tissue supporting the tongue, the gums, and the ligaments and fibers that bind tooth to the socket. Hard tissues refer to the tooth and root. Major advantages of lasers in this field include the ability to induce hemostasis during the procedure, and lower postoperative pain experienced, compared to the more traditional procedures. Dental applications for which lasers are used include gingivectomy (removing excess gum tissue), frenectomy (removal of frenula in the

mouth), and removal of tumors. Lasers that are commonly used for dental procedures are the CO_2, Nd:YAG, argon ion, and holmium:YAG lasers. The collateral thermal damage that results when a CO_2 laser pulse of high peak power and short pulse duration is used on oral soft tissue is in the range 15–170 μm.

The majority of the beam power is absorbed within the first 0.3 mm of soft tissue. The argon ion laser beam, on the other hand, is absorbed in a distance of 1–2 mm, with a collateral damage that is slightly greater than that of the CO_2 laser. The Nd:YAG laser tends to penetrate a lot deeper into soft tissue. On the other hand, in areas where bleeding can be significant, the Nd:YAG, argon ion, and holmium:YAG lasers are the ones to consider.

In using a laser for gingivectomy, an important issue is to prevent damage to underlying bone and tooth substrate. The CO_2 laser is effectively absorbed by watery tissue. Thus, it is the tool of choice for this procedure, since it does not penetrate deeply into the soft tissue and therefore has minimal impact on the underlying bone and tooth structures.

Different forms of oral lesion can be effectively removed using lasers. Oral leukoplakia is one common type of precancerous lesion that lends itself well to laser application. It forms as thick, white patches of the mucous membrane on the tongue, gums, or the inside of the cheeks and cannot be removed simply by scraping. It is normally caused by chronic irritations such as tobacco or alcohol abuse.

The procedure to remove the lesion may involve either excision or vaporization. In either case, no suturing is required. When the lesion is thick and white, then due to its paucity of water, it might be preferable to use excision. Also, since vaporization normally results in superficial layer removal, deeper layers may not be removed, enabling the lesion to recur. For the thin patches that lend themselves to vaporization, a defocused CO_2 laser beam may be used at a power level of 15–20 W. Vaporization normally leaves a thin carbonaceous deposit on the wound. It is not fully established whether or not this is beneficial to the healing process.

Compared to the medical field, the application of lasers in nanotechnology is a relatively new and rapidly growing field. In Section 14.2, we outline some of the developments in nanotechnology that involve the use of lasers.

14.2 NANOTECHNOLOGY APPLICATIONS

Nanotechnology refers to the variety of techniques that are utilized in the creation of features and/or structures with minimum feature dimensions smaller than 100 nm. Several technologies have been developed for fabricating parts with nanosized components and/or features. Our focus in this section, however, will be on laser-based systems. This is an area that is mushrooming, and due to space limitations, we shall only discuss a few examples. Some of the features or structures that can be created using laser-based nanotechnology include nanoholes, grating, nanobumps, nanojets, and nanotubes.

The resolution of features created using laser technology is normally determined by the diffraction limit of the laser system, and that is of the order of half the wavelength of the radiation. However, with ultrashort or femtosecond pulse lasers, special techniques can be used to generate sub-wavelength features. These include selecting the peak laser fluence to be just above the threshold ablation fluence (Figure 9.25), using interferometry, using nanoparticles, and using the laser pulses in combination with an atomic force microscope or a scanning near-field optical microscope. In general, the feature resolution associated with femtosecond lasers can be approximated by:

$$d_r = \frac{k_n \lambda}{\mathrm{NA} q_e^{1/2}} \tag{14.1}$$

where d_r is the femtosecond laser feature resolution, $k_n = 0.5–1$ is a constant, NA is the numerical aperture of the focusing optics, and q_e is the number of photons required to overcome the energy bandgap.

14.2.1 Nanoholes and Grating

A matrix of nanoholes can be generated by ablating a thin film using four interfering femtosecond beams. Using a Ti:sapphire femtosecond laser (pulse duration, 90 fs; wavelength, 800 nm; pulse repetition rate, 1 Hz; beam diameter, 6 mm) with an average laser fluence of 120 mJ/cm^2, a matrix of 800 nm diameter circular holes can be produced in a gold film of thickness 50 nm at intervals of 1.7 μm (Figure 14.2a). Now with two interference beams instead of four, a grating structure is obtained (Figure 14.2b) (see also Section 11.6).

Nanoholes can also be generated by directing the laser beam at the gap between the tip of an atomic force or scanning near-field optical microscope and the thin film. This phenomenon results from electromagnetic field enhancement below the tip and thermal expansion of the tip. When the beam intensity exceeds a threshold value (which is less than the ablation threshold without any particles), each pulse of the laser results in a nanohole being created in the film. The hole size and depth increase with an increase in the beam intensity. When the workpiece is moved linearly at a slow enough speed, a grating structure can be produced.

Finely dispersed nanoholes can also be created by first distributing nanoparticles which are transparent to ultraviolet radiation, for example silica and polystyrene particles with diameters ranging between 150 and 1000 nm, on say, an aluminum film of thickness 35 nm on a silicon substrate. Illuminating the surface with a KrF excimer laser of wavelength of 248 nm, pulse duration of 23 ns, and fluence between 100 and 800 mJ/cm^2 results in an array of nanoholes, with depths and diameters that increase with the laser fluence. Similar holes can also be produced on a silicon substrate using an infrared femtosecond laser after first depositing alumina particles (which are transparent to the infrared radiation) on the silicon substrate. This also results from optical enhancement effect between the particles and substrate.

14.2.2 Nanobumps

If the laser fluence is reduced to about 77 mJ/cm^2, a matrix of conical bumps are formed at locations of maximum interference of the four beams (Figure 14.3a). The size of each bump grows with

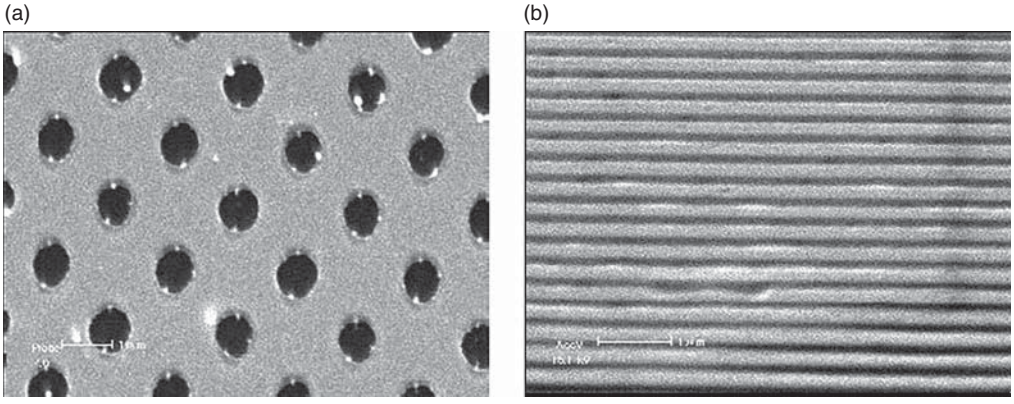


Figure 14.2 (a) A matrix of nanoholes generated using four femtosecond laser beams. (b) A grating generated using two femtosecond laser beams. Source: Nakata et al. (2004)/SPIE.

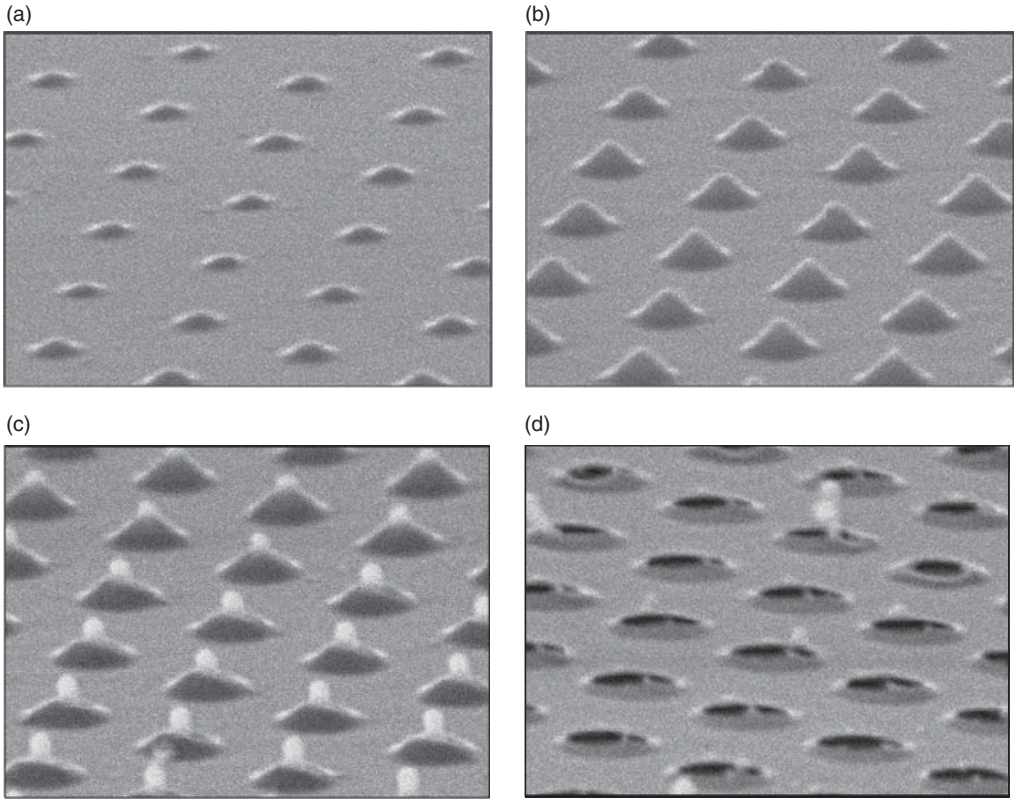


Figure 14.3 Nano-features on a gold film irradiated with four interfering femtosecond lasers. (a) A nanobump matrix at a fluence of 77 mJ/cm^2. (b) The nanobump matrix at a fluence of 89 mJ/cm^2. (c) Nanojets forming from the bumps at a fluence of 97 mJ/cm^2. (d) Nanoholes forming from the bumps at a fluence of 114 mJ/cm^2. Source: Nakata et al. (2004)/SPIE.

increasing fluence (Figure 14.3b). At an even higher fluence, a bead or jet begins to form on top of each bump (Figure 14.3c), and that grows with further increases in the fluence, until eventually the entire bump becomes a nanohole (Figure 14.3d). The hole diameter increases with the laser pulse energy. The minimum bump size obtainable is of the order of 8 nm height and 330 nm diameter.

These are much smaller than the wavelength of the laser used. However, the typical bump aspect ratio (ratio of bump height to diameter) is about 0.44, which is about two orders of magnitude greater than that obtained when nanosecond lasers are used for laser texturing. This is because of the shorter thermal diffusion length (of the order of tens of nanometers) for a femtosecond laser compared to that of the nanosecond laser (of the order of tens of microns). Each of the nano-features, from the bump to the hole formation, has a well-defined threshold. For some materials, for example 100–200 nm chromium-coated layers, the hole may be surrounded by a molten ring along with some droplets, in that region of the material where the beam intensity falls between the melting and ablation thresholds.

The bumps result from vapor pressure that is generated as part of the film at the locations of maximum interference. This peels and expands material that has been softened by heating at those locations. That the material is heated at those locations is supported by the conical shape of the bumps, indicating that the film is more elongated at the top of the bump than at the base. This is because

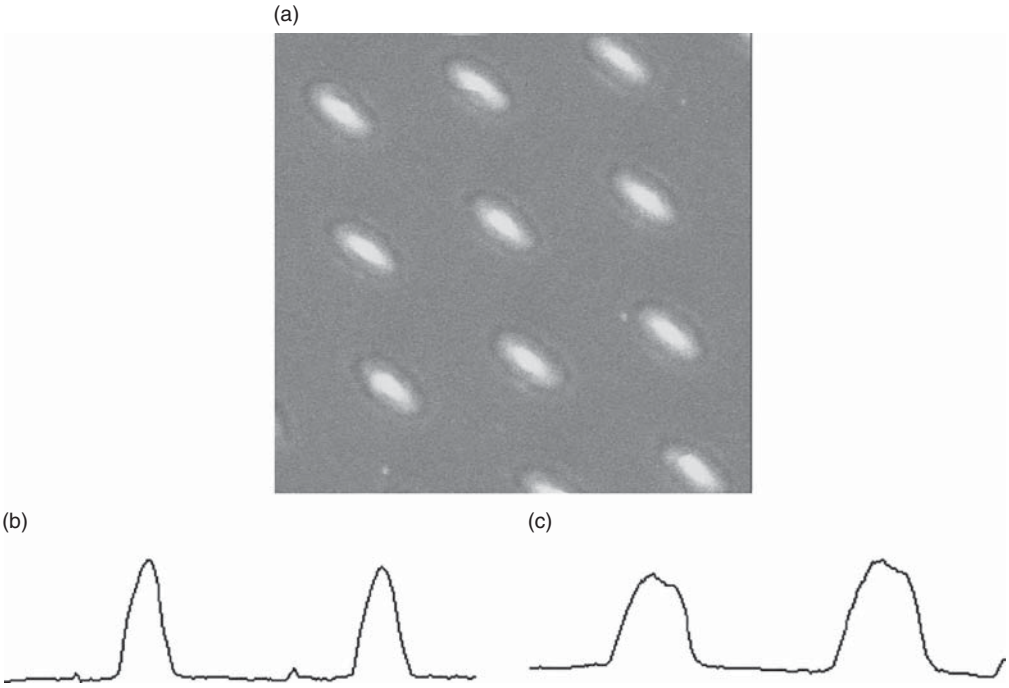


Figure 14.4 Ellipsoidal bumps formed from three beams. (a) Top view. (b) Atomic force microscope scan of a section through the minor axis. (c) Atomic force microscope scan of a section through the major axis. Source: Nakata et al. (2004)/SPIE.

the film is at a higher temperature and therefore more ductile and thin at the top. The formation of the jet further supports this concept and also indicates that the area where the jet forms may have been molten.

The bump structure that is obtained depends on the number of interfering laser beams. Ellipsoidal bumps are formed for three beams (Figure 14.4a–c), and linear bumps are obtained for two beams (Figure 14.5a,b).

14.2.3 Laser-Assisted Nanoimprint Lithography

In traditional nanoimprint lithography (NIL), a mold (often a quartz template) is used to imprint features as small as 20 nm into a low-viscosity material, usually either a thermoplastic or UV-curable resist. Both two- and three-dimensional nanostructures can be produced with a single resist layer. The UV-based process does not require heating. However, multiple resist layers and multiple etching steps are often necessary, making it relatively expensive. The thermal-based process, on the other hand, requires heating of the polymer to soften it. This is a relatively slow process, and it creates a thermal expansion difference between the mold and substrate which could result in misalignment.

Laser-assisted nanoimprint (LAN) lithography uses a laser beam which is transparent to the quartz mold (since the quartz material bandgap is greater than the photon energy), for example, a 20 ns pulse of an XeCl excimer laser of wavelength 308 nm. The beam is directed at the polymer film on the substrate, and at the same time, the mold is imprinted into the polymer. The entire process can take less than 500 ns, since only a single laser pulse is necessary. Lower imprint pressures are required since the laser beam reduces the viscosity of the polymer.

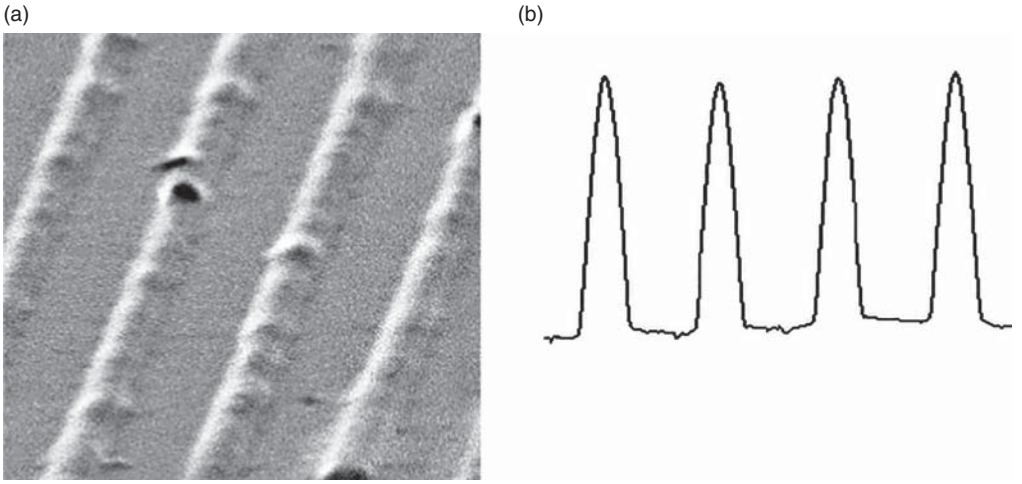


Figure 14.5 Linear bumps formed from two beams. (a) Pictorial view. (b) Atomic force microscope scan of a transverse section. Source: Nakata et al. (2004)/SPIE.

14.3 SUMMARY

In the medical field, lasers are used in the fabrication of medical devices such as stents in therapeutic applications and in medical diagnostics. In stent manufacturing, a laser is used to cut the stent geometry in a metal tubing, followed by finishing operations. The major therapeutic applications include surgery, where they have been used in: LAUP; opthalmology, with applications in diabetic retinopathy to seal leaking blood vessels in the retina; dermatology, for skin resurfacing to improve skin appearance, and also for treating skin tumors; and dentistry, where they are mainly used for soft tissue therapies such as gingivectomy (removing excess gum tissue), and tumor removal. In many of these applications, lasers can be used to minimize hemorrhaging through coagulation and have the advantage of process precision, minimal damage to collateral tissue, and rapid and relatively painless healing.

Laser-based nanotechnology applications include the production of features or structures such as nanoholes, grating, nanobumps, nanojets, and nanotubes. These can be produced in a number of different ways, such as selecting the peak laser fluence to be just above the threshold ablation fluence, using interferometry, using nanoparticles, and using the laser pulses in combination with an atomic force microscope or a scanning near-field optical microscope.

Finally, in LAN, a laser beam is used to heat and soften a polymer film before it is imprinted by a mold.

BIBLIOGRAPHY

Medical References

Berlien, H.P. and Muller, G.J. (2003). *Applied Laser Medicine*. Berlin: Springer-Verlag.

Catone, G. (1994). Laser technology in oral and maxillofacial surgery, Part II: Applications. *Selected Readings in Oral and Maxillofacial Surgery* 3 (5): 1–35.

Constable, I.J. and Lim, A.M. (1990). *Laser – Its Clinical Uses in Eye Diseases*. Edinburgh, UK: Churchill Livingstone.

Peng, Q., Juzeniene, A., Chen, J. et al. (2008). Lasers in medicine. *Reports on Progress in Physics* 71: 1–28.

Singerman, L.J. and Coscas, G. (1999). *Ophthalmic Laser Surgery*, 3e. Boston, MA: Butterworth Heinemann.

Visuri, S.R. (1996). *Laser Irradiation of Dental Hard Tissues*. Doctoral Dissertation. Northwestern University, Evanston, IL.

Nanotechnology References

Hong, M.H., Huang, S.M., Luḱyanchuk, B.S. et al. (2003). Laser assisted nanofabrication. *Proceedings of SPIE* 4977: 142–155.

Korte, F., Serbin, J., Koch, J. et al. (2003). Towards nanostructuring with femtosecond laser pulses. *Journal of Applied Physics A* 77: 229–235.

Korte, F., Koch, J., Serbin, J. et al. (2004). Three-dimensional nanostructuring with femtosecond laser pulses. *IEEE Transactions on Nanotechnology* 3 (4): 468–472.

Lei, S., Zhao, X., Yu, X. et al. (2020). Ultrafast laser applications in manufacturing processes: a state of art review. *ASME Journal of Manufacturing Science and Engineering* 142: 031005. https://doi.org/10.1115/1.4045969.

Lu, Y.F., Zhang, L., Song, W.D. et al. (2002). Particle enhanced near-field optical effect and laser writing for nanostructure fabrication. *Proceedings of SPIE* 4426: 143–145.

Nakata, Y., Okada, T., and Maeda, M. (2004). Generation of uniformly spaced and nano sized sructures by interfered femtosecond laser beams. *Proceedings of SPIE* 5339: 9–19.

Takada, H., Kamata, M., Hagiwara, Y., and Obara, M. (2004). Nanostructure fabrication by femtosecond laser with near-field optical enhancement effect. *Proceedings of SPIE* 5448: 765–772.

Xia, Q., Keimel, C., Ge, H. et al. (2003). Ultrafast patterning of nanostructures in polymers using laser assisted nanoimprint lithography. *Applied Physics Letters* 83 (21): 4417–4419.

15 Sensors for Process Monitoring

Every process has inputs and outputs. The inputs are those parameters that can be changed and collectively used to control the process. The outputs are those parameters that give a measure of the state of the process and its quality. The inputs for laser processing include the beam power, beam size (focusing), and traverse velocity (welding or cutting speed). Others are gas flow rate, gas type (shielding gas, assist gas, and suppressant gas), orientation, workpiece and filler material properties, properties of optics, beam mode structure, and joint configuration.

Primary outputs that provide a direct measure of the process state are usually difficult to measure in real time, that is, while the process is in progress, and without destroying the part. Secondary outputs are more easily measured in real time, but not after the process. The primary outputs for laser welding, for example, include the penetration, bead width, reinforcement (collectively the bead geometry), heat-affected zone size, microstructure, mechanical properties (hardness and strength), residual stresses, and defects (cracks, porosity, undercut, concavity, etc.). The secondary outputs include the temperatures (temperature distribution, peak temperatures, and cooling rate), acoustic emission (AE), electromagnetic radiation, and weld pool motion. In the case of laser cutting, the primary outputs include kerf width, cut edge squareness, microstructure, dross appearance, and surface roughness (striations). Secondary outputs of the laser cutting process include spark discharge beneath the cut (angle and intensity).

To ensure process quality, it is necessary to measure the outputs of interest, and sometimes where necessary, the inputs as well. The process outputs, however, often need special sensors. Ideally, the sensors have to be nonintrusive, rugged enough for the hostile environment, repeatable and reliable, and relatively inexpensive.

In the rest of the chapter, some of the sensors and sensing techniques that are commonly used or are under development for monitoring and diagnostics during laser processing are discussed. The discussion starts with monitoring of the laser beam itself. This is followed by monitoring of the process, with some of the systems that are discussed also being applicable to seam tracking.

15.1 LASER BEAM MONITORING

In this section, we discuss sensors that are used exclusively for monitoring and control of the laser beam. The principal beam characteristics that are often measured include the beam power, propagation (beam size and location), and mode structure. For a continuous wave (CW) laser, the power level is of interest, while a pulsed laser is characterized by such parameters as beam energy, peak power, repetition rate, pulse duration, pulse shape, and average power. Beam propagation is characterized by the variation of diameter with distance (beam divergence) and beam center direction (pointing stability). The mode structure describes how power is distributed across a plane perpendicular to the

Principles of Laser Materials Processing: Developments and Applications, Second Edition. Elijah Kannatey-Asibu, Jr.
© 2023 John Wiley & Sons, Inc. Published 2023 by John Wiley & Sons, Inc.
Companion Website: www.wiley.com/go/Kannatey-Asibu/PrinicplesLaserMaterialsProcessing

propagation direction. Various techniques have been developed for monitoring each of these characteristics, and we shall briefly outline the principal ones in the following Sections 15.1.1 to 15.1.3.

15.1.1 Beam Power

Drifts in the resonator and other disturbances often introduce fluctuations in the laser output power. Common techniques for measuring the output beam power are the pyroelectric detector and the standard beam dump.

15.1.1.1 Pyroelectric or Thermopile Detector This power measuring technique uses a beam splitter to divert a portion of the beam, which is sampled using a chopper. The resulting intermittent signal is then focused onto either a pyroelectric detector or a thermopile that measures the power level. Pyroelectric detectors are further discussed in Section 15.2.4.1.

15.1.1.2 Beam Dump High-power lasers normally have a beam dump as a standard component, and its purpose is to dispose of the energy generated when the laser is idling. The beam dump may also be used as a calorimeter for measuring the output power of the laser. One such device consists of a conical calorimeter through which water is circulated (Figure 15.1). With this device, the beam is intercepted by a water-cooled polished copper substrate and reflected into the cone. Thermocouples mounted at the inlet and outlet of the calorimeter enable the temperature rise in the water as it flows through the calorimeter to be determined. The temperature rise, ΔT, when combined with the flow rate, $\dot{m}$, as determined by a flow meter, enables the beam power q to be determined:

$$q = \dot{m} c_{\mathrm{p}} \Delta T \tag{15.1}$$

One major disadvantage of this device is that the entire beam is absorbed by the beam dump, during which time the beam is not available for processing. It is thus not suitable for in-process monitoring. Furthermore, the response time is slow (with a time constant of the order of 10 seconds) and is also insensitive to small fluctuations in power.

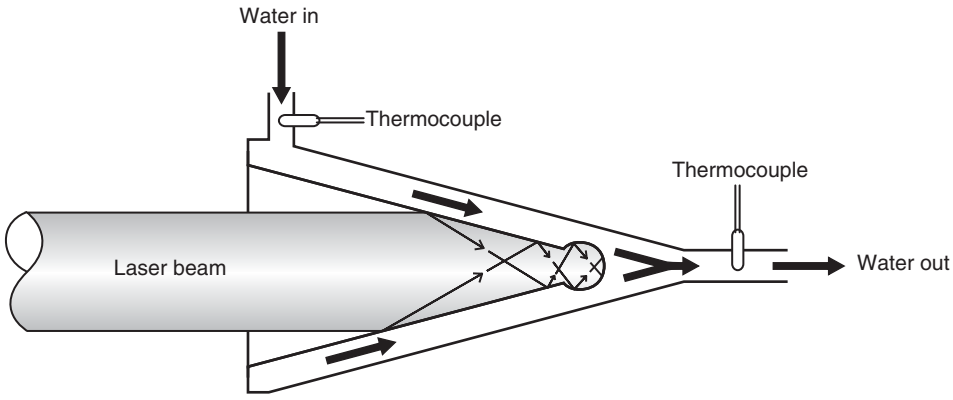


Figure 15.1 Schematic of a cone calorimeter beam dump. Source: From Steen and Mazumder (2010)/with permission of Springer Nature.

Example 15.1 The inlet temperature of a cone calorimeter beam dump is measured to be 25 °C, with the corresponding outlet temperature being 90 °C. Determine the power of the laser beam incident on the calorimeter if the flow rate of water through it is 18 L/h, and the specific heat of water is 4.186 J/g/K = 4186 J/kg/K.

Solution:
From Eq. (15.1), the power is given by

$$q = \dot{m} c_{\mathrm{p}} \Delta T$$

The density of water is 1 g/cm^3 = 1 kg/L. The flow rate is therefore 18 kg/h = 0.005 kg/s.
The power is thus

$$q = \dot{m} c_{\mathrm{p}} \Delta T = 0.005 \times 4186 \times (90 - 25) = 1360 \ \mathrm{W} = 1.36 \ \mathrm{kW}.$$

15.1.2 Beam Mode

The profile or mode of a laser beam can be measured using several techniques:

1. Mechanical methods.
2. Camera-based methods.
3. Approximate methods.

In addition to beam mode (intensity distribution), some of the methods can also provide information on the overall power and beam size (measured as the radius at which the beam power drops to $1/e^2$ of its peak (center) value). These are discussed in the following Sections 15.1.2.1 to 15.1.2.3.

15.1.2.1 Mechanical Methods The mechanical methods are among the earlier techniques developed for beam profiling. The device is normally mounted between the output window and shutter assembly, enabling power monitoring to continue even when the laser is idling. They can also be used to evaluate the beam profile at the focal point. A complete mapping of the beam profile may take about five seconds. The power loss associated with this technique is only about 0.1%. The response time for the device can be as small as 10 ms. Another advantage of the mechanical methods is their high resolution, of the order of 1 μm. Furthermore, they can be used directly with high-power lasers without attenuation since only a minute fraction of the beam is incident on the detector at any instance. However, they are only applicable to CW lasers, since synchronization with pulsed beams can be problematic.

The three most common techniques developed are the rotating rod, rotating hollow needle, and knife edge (or slit).

Rotating Rod Technique This technique uses a rotating circular rod that is reflective. The rod is usually made of molybdenum and is rapidly swept through the laser beam in a plane perpendicular to the beam direction (Figure 15.2). Light is reflected from the rod surface in all directions as it sweeps through the beam. A small portion of the beam light is reflected from the rod onto two pyroelectric detectors that are located such that they measure light that are reflected in orthogonal directions. The detectors may have germanium-coated windows that will only transmit radiation of a specific wavelength. The output voltage of each pyroelectric detector is proportional to the laser power incident on it.

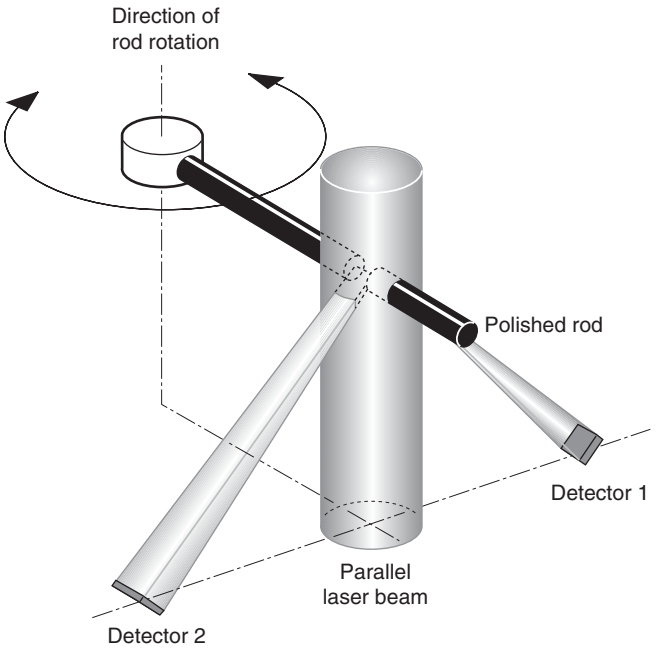


Figure 15.2 Principle of the rotating rod laser beam analyzer. Source: Adapted from Lim and Steen (1984).

Rotating Hollow Needle Technique With this technique, a long slender needle that is hollow inside has a pinhole drilled close to one end of it and is mounted such that the pinhole faces the laser beam (Figure 15.3). Photons that enter the pinhole are deflected to a pyroelectric detector using two mirrors (Figure 15.4). The needle rotation enables the laser beam to be scanned in one direction, say, the *x*-direction. Scanning in the *y*-direction to obtain a two-dimensional image is achieved by

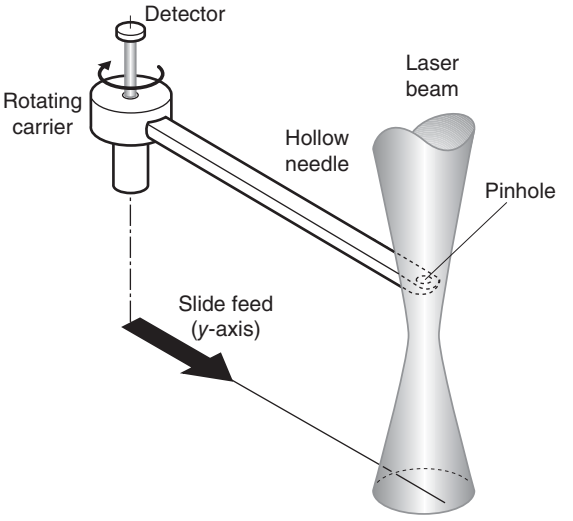


Figure 15.3 Schematic of the rotating hollow needle-based laser beam analyzer. Source: From ICALEO 1992 Proceedings, Vol. 75, Copyright 1993, Laser Institute of America. All rights reserved.

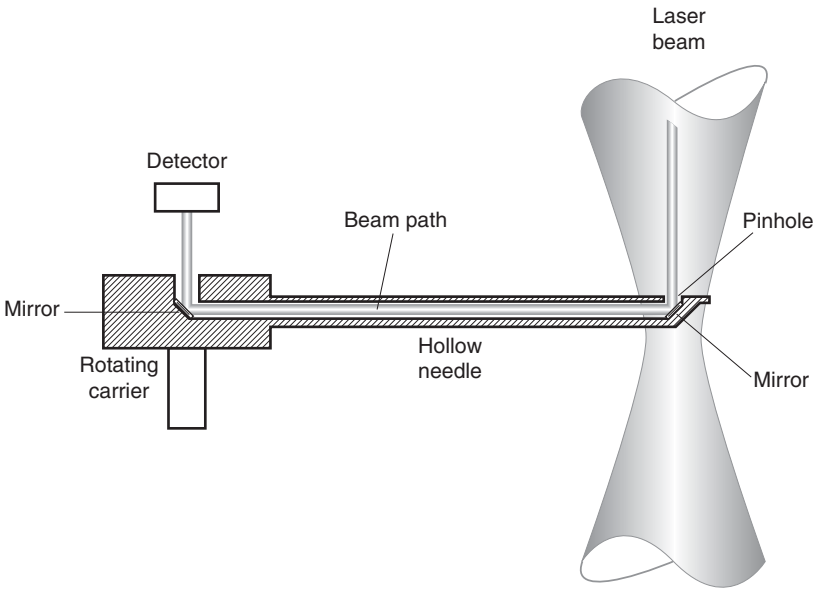


Figure 15.4 Section view of the rotating hollow needle-based laser beam analyzer. Source: From ICALEO 1992 Proceedings, Vol. 75, Copyright 1993, Laser Institute of America. All rights reserved.

Table 15.1 Rotating hollow needle analyzer parameters.

Parameter	Unfocused beam	Focused beam
Pinhole diameter (μm)	100	20–50
Measurement resolution (μm)	75–1500	6–100

moving the axis of rotation in the *y*-direction. The size of pinhole used depends on the beam intensity. Pinhole diameters and measurement resolution are summarized in Table 15.1.

Knife Edge The knife-edge scanning slit profiler is illustrated in Figure 15.5. The laser beam is incident on a rotating drum that has a slit in it. The slit allows a narrow segment of the beam to be incident on the detector at any instant, as the drum rotation enables the slit to scan the beam cross-section.

15.1.2.2 Camera-Based Methods The camera-based methods use sensor arrays and are therefore similar in principle to optical detectors discussed in Section 15.2.5, capable of capturing two-dimensional images of the beam power. Different types of camera are useful for different regions of the electromagnetic spectrum. They have the advantage of being applicable to both CW and pulsed beams. However, the resolution is generally low (about 10 μm) compared to that of the mechanical methods since it is determined by the pixel size.

Two of the most common camera types in use are the charge-coupled device (CCD) and charge injection device (CID). They are most suited for the visible and near infrared wavelength ranges. CID cameras are not appropriate for pulsed beams since it is difficult to synchronize the camera readout with the pulse trigger. The signal-to-noise ratio of CCD cameras may range from 8-bit for the low-end cameras to 16-bit for the high-end cameras.

Other camera types in use include solid-state InGaAs and pyroelectric arrays. However, these are more expensive than the CCD cameras. The pyroelectric cameras are suited for the 1.1–400 μm

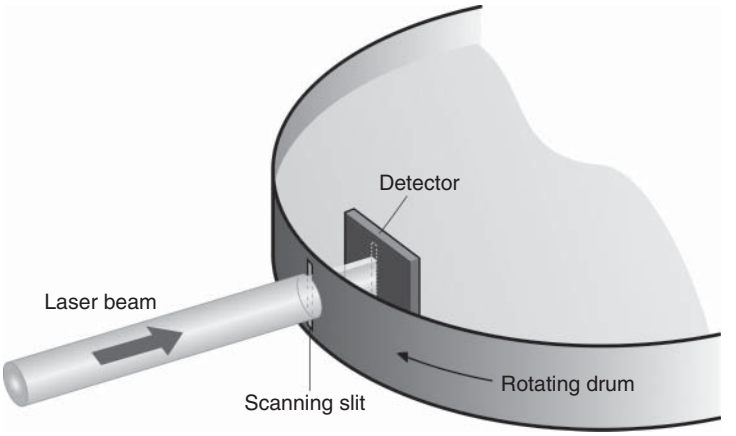


Figure 15.5 Schematic of knife-edge scanning slit laser beam profiler.

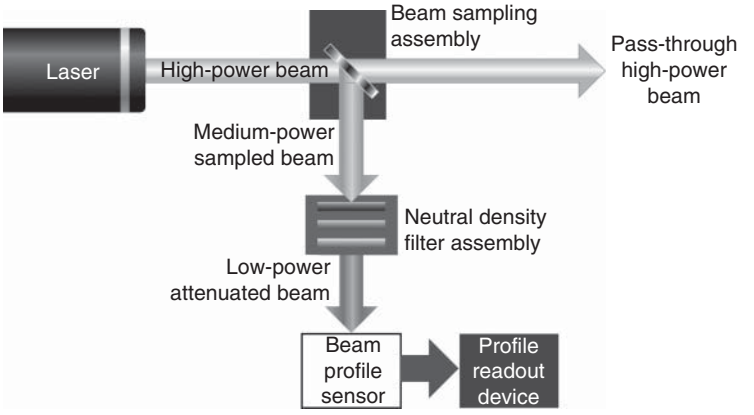


Figure 15.6 Beam attenuation by diversion and/or neutral density filters.

wavelength range, have reliable linear outputs, and are good for pulsed beams. However, CW beams need to be chopped to provide intermittent cooling of the sensor, and the cameras do have lower resolution compared to the CCD and CID cameras.

The camera-based methods are so sensitive to the laser beam that it is necessary to attenuate the beam before it reaches the detector. This is usually done by using neutral density (ND) filters or diverting a small portion of the beam for measurement (Figure 15.6). The attenuation needed for CCD cameras is typically between 10^5 and 10^{12}, while that for pyroelectric cameras is of the order of 10^4. One common type of ND filter used consists of glass impregnated with absorbing material. It has a useful wavelength range of about 0.350–1 μm. The filters are characterized by an ND number defined as

$$ND = \log T \tag{15.2}$$

where T is the transmission ratio of input to output.

Depending on the application, focused laser beams can be as small as 1 μm in size, and since this is smaller than the pixel size of many cameras, it is preferable to use one of the mechanical methods for scanning the focused beam.

In recent years, efforts have been made to use smartphone technology to monitor beam parameters such as size, shape, divergence, M^2 factor, and power.

15.1.2.3 Approximate Methods Plastic burn analysis is the most common form of approximate method used and provides a qualitative measure of beam profiles, especially CO_2 laser beams. The material most commonly used is polymethylmethacrylate (PMMA or acrylic). It is a transparent polymer with a low thermal conductivity and high absorptivity at the CO_2 laser wavelength of 10.6 μm. These properties, in addition to its low sublimation temperature of 300 °C, make it ideal for identification of beam mode. The process involves manually pulsing the laser at low power, about 50 W, until a deep enough cavity is formed in the PMMA material, and measuring the shape formed. The shape that is measured represents an imprint of the beam power distribution, since the depth to which the PMMA is sublimated at any location is determined by the beam power at that location.

15.1.3 Beam Size

The beam size can be measured in a number of ways, but the easiest approach uses a Kapton film.

15.1.3.1 Kapton Film Kapton films are highly suited to measuring the sizes of pulsed beams. They are nonflammable, have low thermal conductivity, and have no melting point. The characteristics of the hole produced in the Kapton film depend on a number of factors, most important being the beam intensity. When the intensity exceeds a threshold value, the film undergoes pyrolytic decomposition to form a carbonaceous char. At yet higher intensities, the film is vaporized. For pulsed beams, the length of time affects the shape of hole produced. Well-defined holes are produced at intermediate pulse times. For very short or very long pulse times, an irregularly- shaped hole is produced. Finally, the size of hole produced depends on the peak pulse power and film thickness.

The hole diameter d_h produced in a Kapton film by a pulsed multimode beam depends on the total pulse energy, Q_e, and is given empirically by

$$d_h = 2w_b - \frac{2\pi w_b^3 h \left[1 - \left(\frac{w_t}{w_b}\right)^3\right] \Delta H_v}{3(1-R)} \left(\frac{1}{Q_e}\right) \tag{15.3}$$

where h is the Kapton film thickness, R is the reflection coefficient, w_b is the base radius of the beam, whose intensity distribution is assumed to be of truncated shape, w_t is the top radius of the beam intensity distribution, and ΔH_v is the energy necessary to heat and vaporize a unit volume of the material.

From Eq. (15.3), it is evident that the hole diameter approaches the beam diameter as the pulse energy approaches infinity, that is,

$$d_h \to 2w_b \quad \text{as} \quad Q_e \to \infty$$

Thus, the beam diameter can be obtained by simply plotting the hole diameter versus the corresponding $1/Q_e$ and extrapolating the resulting line to the diameter axis.

For a Gaussian distributed beam, the hole diameter is given by

$$d_h^2 = 2w^2 \left(\ln Q_e - \ln\left[\frac{\pi w^2 h \Delta H_v}{2(1-R)}\right]\right) \tag{15.4}$$

where w is the $1/e^2$ beam radius.

When the beam radius is constant, the second term of Eq. (15.4) is constant. The square of the hole diameter is then directly proportional to the natural log of Q_e. The beam size can thus be obtained from the slope of the d_h^2 versus ln Q_e plot for a fixed mean power level. The beam size determined by either method is independent of the thickness of the Kapton film, which is typically in the range of 25–125 μm.

15.1.3.2 Other Methods A simple way to determine the beam size is to use a variable aperture to determine the area of the beam that has 87.5% of the total beam energy. The aperture is opened to first measure the total power and then reduced in size until the transmitted power is 87.5% of the total. The output of the aperture is directed toward any of the power measuring devices outlined in Sections 15.1.1 and 15.1.2. The radius of a circle that has the same area is defined as the beam radius. This method is most appropriate if the beam is circular.

Other methods for obtaining an approximate measure of the beam size include burning a hole in a metal foil or photographic paper and measuring the resulting diameter.

15.2 PROCESS MONITORING

The principal parameters that need to be monitored during laser welding, for example, include the weld pool geometry (width and penetration), defects or discontinuities (cracking, porosity, etc.), microstructure (strength), residual stresses, and temperatures. For laser cutting, we have the kerf size, striations, dross, and microstructure. Among the most commonly used sensors for real-time monitoring are AE, audible sound (AS) (acoustic sensing), and infrared/ultraviolet detectors, all of which are secondary outputs.

15.2.1 Acoustic Emission (AE)

AE refers to stress waves that are generated as a result of the rapid release of elastic strain energy within a material due to a rearrangement of its internal structure. It is also sometimes referred to as stress wave emission. The resulting stress waves propagate through the structure and produce small displacements on the surface of the structure. These are detected by sensors that convert the displacements into electrical signals. AE is an active phenomenon, since it is generated by the process under investigation. As a result, AE signals are well suited for real time or continuous monitoring because they are generated while the phenomenon is undergoing change.

15.2.1.1 AE Detection Two types of transducers are normally used for AE signal detection, namely, piezoelectric and capacitive transducers. The piezoelectric transducer is more commonly used.

Piezoelectric Transducers The principle of piezoelectric transducers is based on the characteristics of a damped spring-mass system attached to a frame (Figure 15.7). The signal to be detected subjects the base frame to an input displacement, y, assumed to be sinusoidal and given by

$$y = Y \sin \omega t \tag{15.5}$$

where Y is the maximum amplitude of the input base displacement (m) and ω is the angular frequency of the input base oscillation (rad/s).

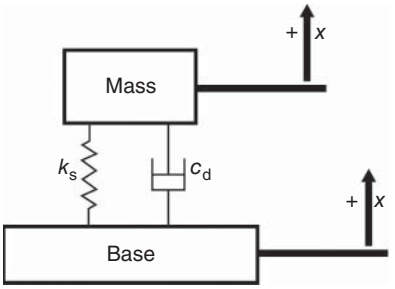


Figure 15.7 Schematic of simple damped spring-mass system.

Assuming that both the mass and base are displaced in the positive direction and that x is greater than y and $\dot{x}$ is greater than $\dot{y}$, then the equation of motion of the system shown in Figure 15.7 is given by

$$m_a\ddot{x} = -k_s(x - y) - c_d(\dot{x} - \dot{y}) \tag{15.6}$$

where c_d is the damping coefficient (N s/m), k_s is the spring constant (N/m), m_a is the system mass (kg), and x is the displacement of the mass (m).

Now let $z = x - y$, the relative displacement between the base and mass. Then it can be shown that

$$\frac{Z}{Y} = \frac{\omega^2}{\omega_n^2}\left[\left(1 - \frac{\omega^2}{\omega_n^2}\right)^2 + \left(2\varsigma\frac{\omega}{\omega_n}\right)^2\right]^{-1/2} \tag{15.7}$$

where Z is the maximum amplitude of the relative motion between the base and mass (m), ω_n is the natural frequency of vibration of the undamped system (rad/s), and ζ is the damping ratio.

From Eq. (15.7), we note that when $\omega_n \ll \omega$, that is, the natural frequency of the device is much lower than the lowest harmonic of the motion to be measured, then Eq. (15.7) approaches unity. Thus,

$$Y \approx Z \quad \text{for} \quad \frac{\omega}{\omega_n} \gg 1 \tag{15.8}$$

In other words, the relative motion between the mass and frame then approaches the absolute motion of the frame. In effect, the mass remains motionless in space. x is the displacement of the mass which becomes quasi-static at the high frequencies, that is $x \approx 0$. The phase difference does not affect the magnitude of the reading. Thus by replacing the spring and damper system with a piezoelectric crystal (Figure 15.8), the relative motion between the base and mass will strain the crystal, generating electrical signals which are a measure of the displacements due to the stress waves.

A material that is commonly used to induce the piezoelectric phenomenon in AE transducers is lead zirconate titanate (PZT) which is a ceramic element, and is most commonly used because it has a relatively high dielectric constant. The relative dielectric constant of a material is the ratio of charge density arising from an electric field with and without the material present.

Capacitive Transducers This type of transducer is based on the principle of the capacitor, where the capacitance between two parallel plates depends on the distance between the plates. The mass in this case is a metallic cylindrical piece placed with its flat surface parallel to the structure surface.

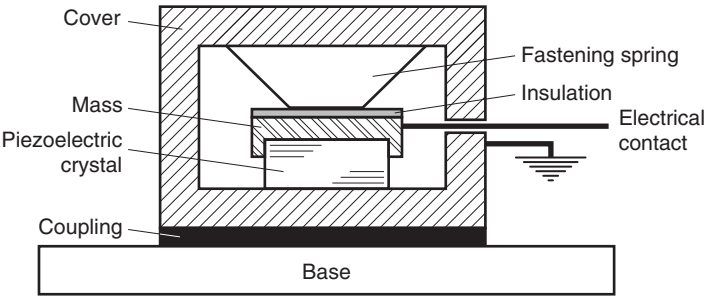


Figure 15.8 Schematic of a piezoelectric transducer.

The space between the mass and structure is statically charged, and changes in the gap result in a change in the capacitance of the system and consequently a change in the voltage across the capacitor.

15.2.1.2 Background This section provides a brief summary of some important factors associated with AE monitoring. Specifically addressed are the following:

Coupling: Since air is a poor transmitter of high frequency waves, the transducer needs to be in very good acoustic contact with the workpiece under inspection. A couplant is normally applied to the transducer base for effective coupling. Typical couplants are oils, glycerin, and greases.

Pre-processing: The AE signals generated are normally of very low amplitude and high frequency content, and need to be amplified close to the signal source using a low-noise pre amplifier, to minimize noise contamination as the signal is propagated over the transmission cable. Most preamplifiers have a gain of 40–60 dB (100–1000 times).

Signal forms: There are two principal types of AE signals:

1. *Burst Signals*: These are typical of individual events, for example, cracking in weldments. An event is a single occurrence of an AE activity.
2. *Continuous Signals*: These are typical of overlapping series of events, for example, continuous welding without defects.

Precautions: Some of the major precautions that need to be taken in applying conventional AE instrumentation include the following:

Protecting the transducer from the high temperatures of the processing environments.
Providing a highly reliable acoustic contact between the transducer and the structure.

AE Sources and Major Uses: The common sources of AE include plastic deformation processes such as dislocation motion, twinning, and grain boundary sliding; phase transformations, such as martensitic transformation; decohesion of inclusions; fracture; and friction. The detectability of an AE signal depends on the energy release rate of the source. Thus slow processes such as creep and most phase transformations do not generate detectable AE. On the other hand, processes such as twinning, martensitic phase transformation, and a short pulse laser beam, that occur over a short period of time, tend to be prolific sources of AE.

Given the sources of AE, it would be most useful for detecting cracking in the fusion zone, and martensitic transformation in both the fusion and heat affected zones. It would also be sensitive to keyhole failure, and to some extent, porosity formation.

15.2.1.3 AE Transmission Stress waves normally propagate in a material in any of four different wave modes, namely, shear (transverse) waves; longitudinal (pressure) waves; Rayleigh (surface) waves; and Lamb waves. The wave type that predominates in a detected signal depends on the locations of the transducer and source, and the structure through which the signal propagates.

15.2.1.4 Traditional AE Signal Analysis Traditional methods used in analyzing AE signals include count and count rate (record of the signal ring-down pulses whose amplitude exceed a preset threshold voltage); amplitude distribution analysis; frequency spectrum; autocorrelation function; root mean square (RMS) value of the signal. Factors that influence the results include the threshold setting (for count and count rate only); gain setting or amplification; and frequency bandwidth.

The threshold setting is a voltage level that is used for the signal counts. A signal count is registered each time the signal level exceeds the threshold setting. The latter is normally set higher than the background noise that could interfere with desired signals. The frequency bandwidth cuts off unwanted higher and/or lower frequency signals, while the gain enables the initial low signal amplitude to be increased. Figure 15.9 illustrates some of the traditional AE signal processing terminology.

A variation of AE detection is what is termed the acoustic mirror. This is further discussed in the following Section 15.2.2.

15.2.2 Acoustic Mirror

The acoustic mirror device arises from the observation that high frequency signals are generated when a high intensity laser beam is incident on a mirror. The signals are detected using a piezoelectric transducer mounted on the back-side of the mirror (Figure 15.10). The detected signals are sensitive to a number of variables, including the beam power, size, its location on the mirror, wavelength,

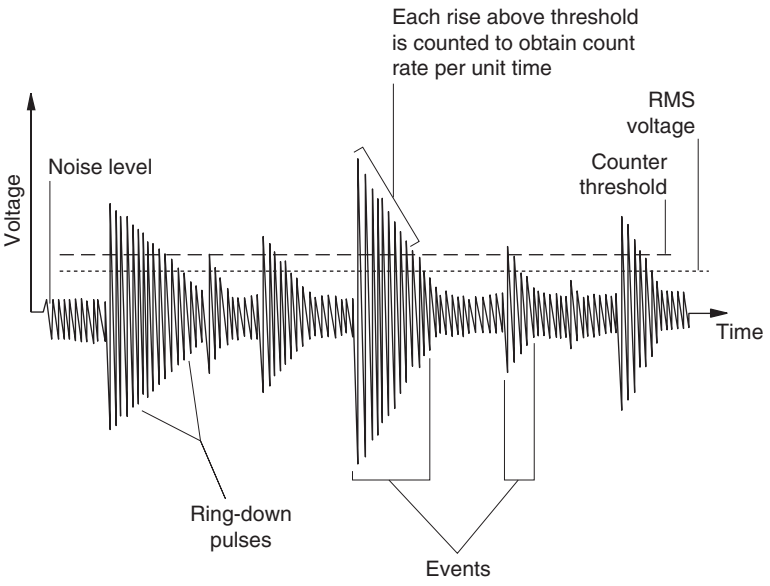


Figure 15.9 Traditional acoustic emission signal processing terminology. Source: From Kannatey-Asibu and Dornfeld (1981)/The American Society of Mechanical Engineers.

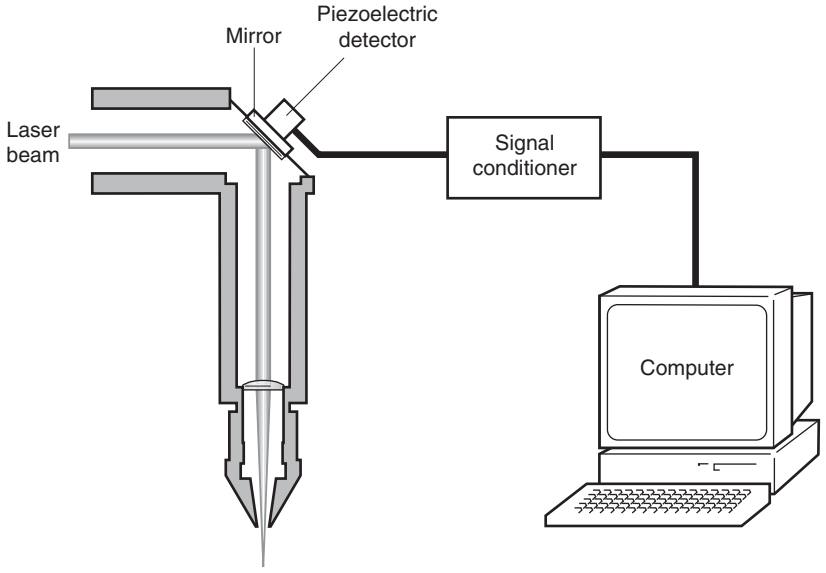


Figure 15.10 An acoustic mirror detector. Source: Adapted from Steen and Weerasinghe (1986).

laser cavity gas composition, and back reflections from the workpiece. The signal intensity increases with laser power, and decreases with beam diameter on the detector mirror. The signals are stress waves that are generated by localized expansion of the mirror when it absorbs a certain percentage of the incident laser radiation that instantaneously heats the mirror surface. The recorded signal is due to power variation, and not the absolute power. A typical signal output is shown in Figure 15.11, illustrating the long term characteristics of the signal, as well as its dependence on keyhole initiation. Spectral analysis of the acoustic mirror signal normally shows two principal peaks, one around 100 kHz, and the other around 1 MHz. The low frequency component is found to depend on the gas

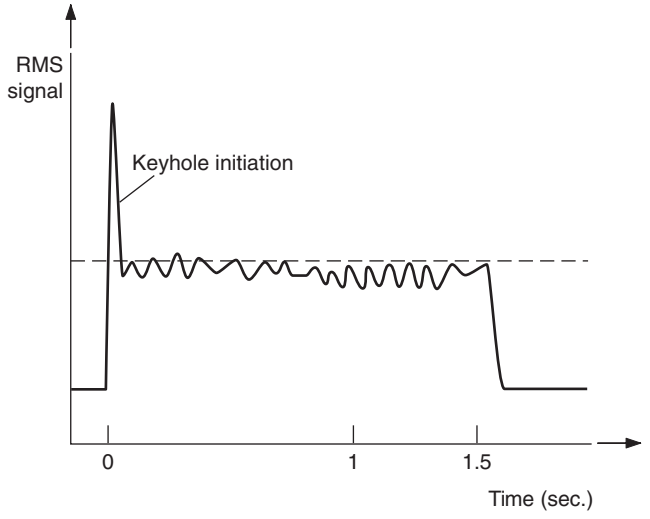


Figure 15.11 An acoustic mirror output signal. Source: Adapted from Steen and Weerasinghe (1986).

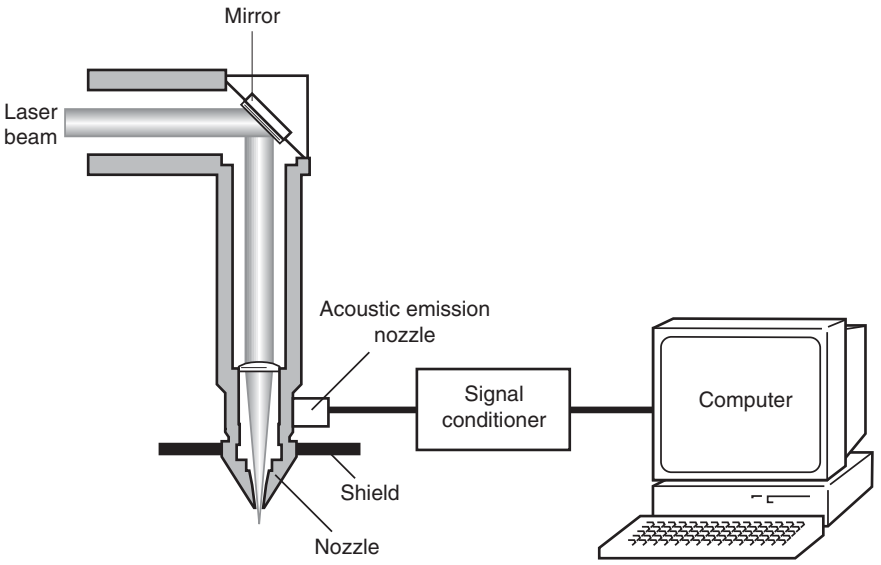


Figure 15.12 Acoustic emission nozzle setup. Source: From *ICALEO 1992 Proceedings*, Vol. 75. Copyright 1993, Laser Institute of America. All rights reserved.

mixture, while the high-frequency component depends on the cavity design, and thus on the cavity oscillations. The dependence of the acoustic mirror signal on beam location on the mirror results from the variation in the distance between the beam location and the sensor location.

A major advantage of the acoustic mirror device is that it does not interfere with the incident laser beam, and is also independent of the processing direction. The mirror can be located anywhere in the beam delivery path.

A variation of this monitoring technique is what is called the AE nozzle (AEN), and involves mounting a piezoelectric transducer on a beam delivery tube which is thermally insulated from the nozzle by a washer. The detector is also protected from direct back reflections from the process by a shield (Figure 15.12). The characteristics of the signal detected under these conditions are found to be a good indication of the cut quality in laser cutting (Figure 15.13).

15.2.3 Audible Sound (AS) Emission

The AS technique uses a microphone to detect low frequency (15 Hz to 20 kHz) acoustic signals generated during laser processing, in a sense, making use of the sound detected by an operator. The acoustic signals are the result of pressure variations at the process source, which propagate to the microphone. The intensity of the detected signal thus depends on the pressure and velocity of the fluid at the source. It also depends on the size of the source, and the distance between the source and sensor. In laser cutting, the signal intensity depends on the density and velocity of the assist gas jet, and the resonant frequency of the detected signal is found to depend on the cutting front geometry, specifically the kerf width and depth of cut. The signals associated with gas flow rates up to 40 L/min and velocities of the order of 50 m/s in laser welding are barely measurable. The microphone is normally mounted in the immediate vicinity (approximately, 40 mm) and directed toward the process area (at an angle of about 45°). It has the advantage of being non-contact. A typical setup is illustrated in Figure 15.14.

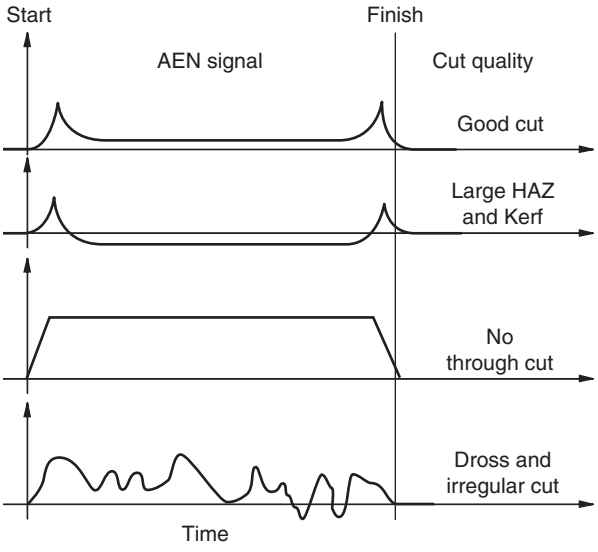


Figure 15.13 Acoustic emission nozzle response during laser cutting. Source: From *ICALEO 1992 Proceedings*, Vol. 75. Copyright 1993, Laser Institute of America. All rights reserved.

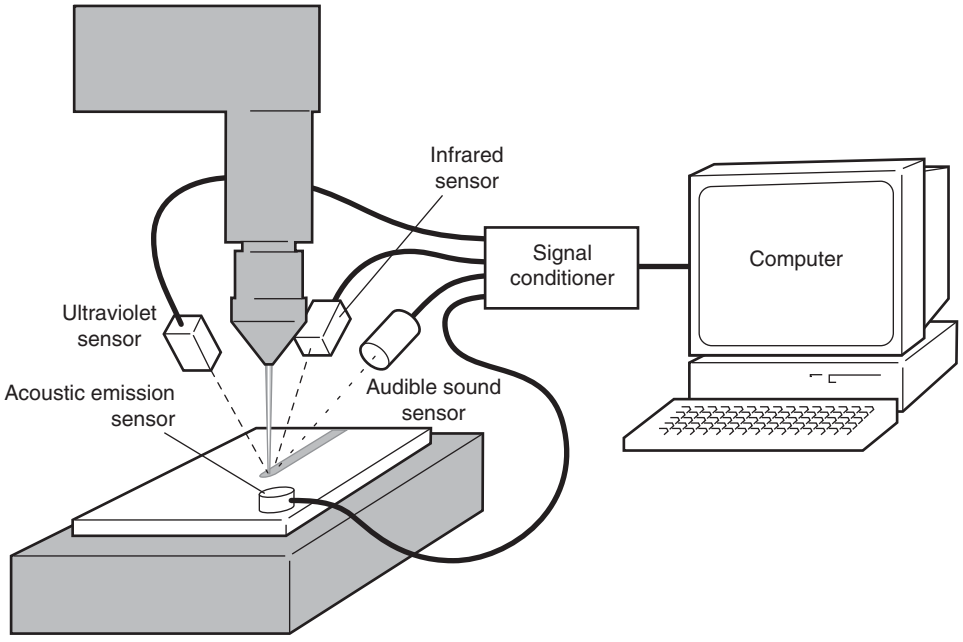


Figure 15.14 Noncontact detection system for audible sound, infrared, and ultraviolet signals. Source: From *Journal of Laser Applications*, Vol. 14, No. 2. Copyright 2002, Laser Institute of America. All rights reserved.

Sources of sound emission during laser processing include the following:

1. Variations in gas momentum caused by heating and boiling processes (vapor formation).
2. Variations in gas pressure or momentum change associated with the impact of gas jets on the workpiece surface. This is especially important in laser cutting where the signals may result from vibrations caused by impingment of the assist gas on the cutting front.
3. Vapor expansion out of the keyhole opening in laser welding. The contribution of this signal source depends on
   i. Vapor pressure inside the keyhole.
   ii. Size of the keyhole.
   iii. Velocity of the expanding front.
4. Shock wave accompanying plasma ignition, which would result in a transient signal. However, when the plasma obstructs the laser beam and prevents it from reaching the workpiece, then the intensity of the subsequent signal is significantly reduced.
5. Defects such as cracking.

15.2.4 Infrared/Ultraviolet (IR/UV) Detection Techniques

In this section, general information is first presented on Infrared/Ultraviolet (IR/UV) detection, and how that relates to laser processing. This is then followed by more detailed discussion of infrared detection, and then ultraviolet (UV) detection.

IR/UV detection techniques analyze radiation emitted from the process zone in two wavelength bands of the electromagnetic spectrum. One is the infrared band in which most of the radiation from the hot material is concentrated, and the other is the ultraviolet band in which the plasma radiation is concentrated. Changes in the temperature of the plasma emerging from the keyhole will result in a change in the ultraviolet signal, which can be used to provide a predictive indication of keyhole failure. This is effective because the response of liquid flow is much slower than that of the plasma. For example, should there be a sudden reduction in the plasma density, that could reduce the effectiveness of beam coupling to the keyhole, resulting in possible eventual failure of the keyhole and the welding process. Thus by detecting the change in the ultraviolet signal, the impending failure can be anticipated. The intensity of the ultraviolet signal is significantly reduced when the plasma is blown through the workpiece to form a cut, further confirming the dependence of the ultraviolet signal on the plasma. Changes in the infrared signal also provide information on variations in the melt pool.

A typical sensor used for the infrared radiation is a germanium photodiode fitted with a silicon filter having a spectral range from 1.0 to 1.9 μm. The ultraviolet radiation may be measured with a gallium phosphide (GaP) photodiode with a spectral range from 0.19 to 0.52 μm. Possible arrangements for the set-up would be as shown in Figure 15.14. Both the ultraviolet and infrared signal intensities increase with the laser power, while increasing shielding gas flow rate reduces the signal intensities, probably due to a reduction in plasma volume. Table 15.2 illustrates how the relative intensities of the infrared and ultraviolet signals reflect on weld quality.

15.2.4.1 Infrared Detection

Background At temperatures above absolute zero, all objects radiate infrared energy that is part of the electromagnetic spectrum and has a wavelength range from 0.75 to 10^3 μm (between the visible light and microwave bands). The bodies that emit infrared radiation can be classified into three types based on the spectral characteristics of the radiation:

Table 15.2 Weld quality and its dependence on infrared/ultraviolet signals.

Ultraviolet signal	Infrared signal	Weld quality
High	Low	Good
High	High	Poor
Low	High	Cut
Low	Low	Hole, low penetration
Oscillation	Oscillation	Hump and perforation

Source: Chen et al. (1991).

1. *Blackbody Radiation*: This type of radiation is emitted from a body with an emissivity of one. A blackbody is one that absorbs all incident radiation and reflects none.
2. *Graybody Radiation*: The emissivity distribution of this type of body is similar to that of a blackbody, but less than unity.
3. *Selective Radiation*: In this case, the emissivity varies with wavelength.

Emissivity is the ratio of the radiant energy emitted by a surface to that emitted by a blackbody at the same temperature. Also, the fraction of an incident beam of radiant power that is absorbed by a surface. It usually depends on the direction of the incident radiation and is greatest in the normal direction and declines in magnitude in proportion to the cosine of the angle with the normal. The intensity of the radiated power (W/mm^2) depends on both the temperature of the body and the wavelength. For a blackbody, the distribution is given by Planck's law, which can also be expressed as

$$I_\lambda = \frac{c_1 \lambda^{-5}}{e^{\frac{c_2}{\lambda T}} - 1} \tag{15.9}$$

where $c_1 = 2\pi {c_0}^2 h_\mathrm{p} = 3.74 \times 10^{-16}\ \mathrm{J\,m^2/s}$, $c_2 = c_0 h_\mathrm{p}/k_\mathrm{B} = 1.44 \times 10^{-2}\ \mathrm{m\,K}$, and I_λ is the amplitude of radiant energy at a given wavelength λ, at temperature T, per unit wavelength interval, per unit time, per unit solid angle, per unit area.

The energy distributions for a blackbody are shown in Figure 15.15a,b on log–log and linear plots, respectively, and those for a graybody and non-graybody (where radiation is selective) are illustrated in Figure 15.16. For a blackbody, each temperature results in a specific curve that increases rapidly from zero intensity at zero wavelength to a maximum and then gradually decreases asymptotically toward zero at long wavelengths. The peak intensity increases with temperature and occurs at shorter wavelengths for higher temperatures. The wavelength λ_m at which the peak intensity occurs for a given temperature is given by Wien's law:

$$\lambda_\mathrm{m} T = \text{constant} \tag{15.10}$$

For a blackbody at a given temperature, T, the total hemispherical radiation I per unit area and time is given by the area under the distribution curve for that temperature, and is obtained by integrating Planck's equation. That results in the Stefan–Boltzmann law:

$$I = \int_0^\infty I_\lambda d\lambda = \sigma_\mathrm{B}(T^4 - T_0^4) \tag{15.11}$$

where σ_B is the Stefan–Boltzmann constant $= 5.6704 \times 10^{-8}\ \mathrm{W/m^2\ K^4}$.

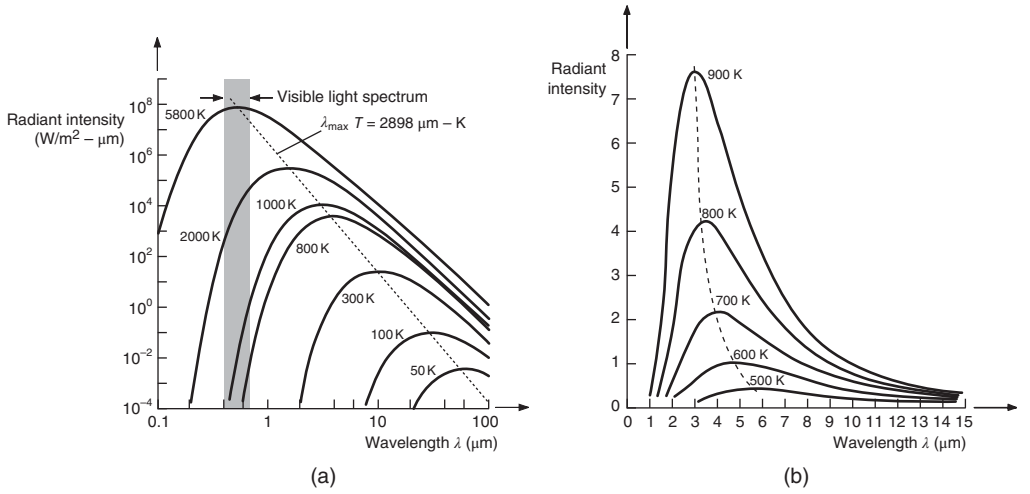


Figure 15.15 Energy distribution for (a) blackbody on a log scale and (b) blackbody on a linear scale. Source: From Benedict (1984).

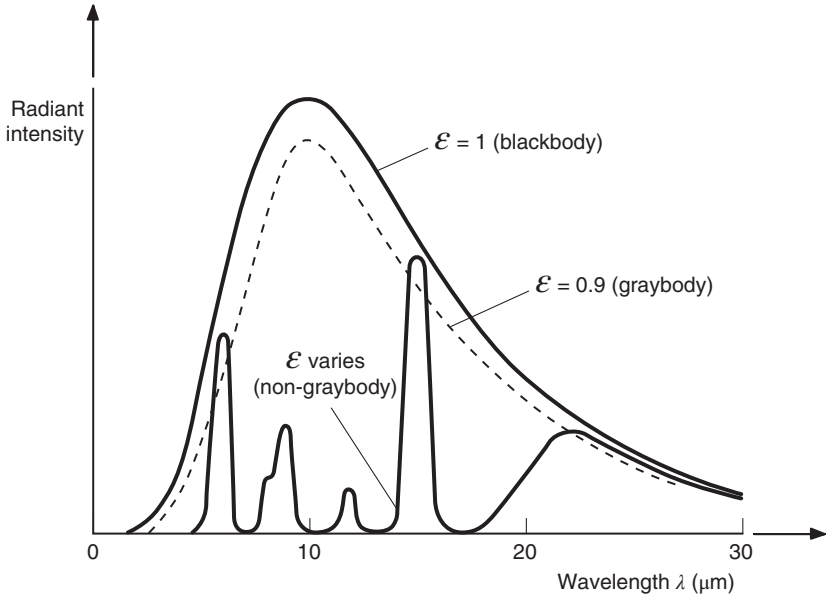


Figure 15.16 Energy distribution for a graybody and a non-graybody. Source: From Benedict (1984).

For a graybody, the total radiation is the product of that of the blackbody and the emissivity of the body in question. The emissivity generally depends on the material and its surface conditions. The temperature of a graybody whose emissivity and ambient temperature is known can be determined by measuring the radiance and using the Stefan–Boltzmann law.

Infrared Detectors The sensors that are used for detecting infrared radiation can be categorized into two broad groups:

1. *Thermal Detectors*: With this type of detector, the material absorbs radiation, resulting in the generation of phonons that cause lattice heating. The lattice heating produces a change in the electrical properties of the material that are detected. The transduction process thus involves two steps. Common types of thermal detectors include pyroelectric devices and thermocouples.
   (a) *Pyroelectric Devices*: A change in temperature due to the incident radiation results in a change in dipole moment of the material, and thus charge difference between crystal faces, which is measured. They have fast response, but require chopped radiation source to define absolute levels.
   (b) *Thermocouples*: The principle of thermocouples is based on the Seebeck effect, where thermal energy is converted into electric energy when two dissimilar metals are joined and one end is heated. If the joined wires form a closed circuit, in which case we also have two junctions, then electromotive force (EMF) is induced and results in current flow through the circuit. One of the junctions, the hot junction, is exposed to the incident radiation, while the cold junction remains at a relatively constant temperature. The cold junction is usually located near massive portions of the mounting or case.

   For two dissimilar metals, A and B, joined as shown in Figure 15.17, the net induced EMF (E_v) is given by

$$E_v = \int_{T_1}^{T_2} \alpha_A dT + \int_{T_2}^{T_1} \alpha_B dT$$
$$= \int_{T_1}^{T_2} (\alpha_A - \alpha_B) dT \tag{15.12}$$

   where α_A and α_B are the absolute Seebeck coefficients of materials A and B, respectively (V/K).

   The voltage, E_v, is thus used as a measure of the temperature difference between the hot and cold junctions.
2. *Photon Detectors*: With photon detectors, the radiation absorbed directly produces a change in the electrical properties of the detector and is thus one-step. Photon detectors thus have a faster response, and the examples are photoconductive, photoresistive, photoemissive, and photovoltaic detectors.

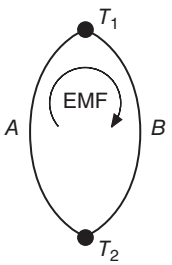


Figure 15.17 Basic schematic of a thermocouple junction.

Infrared Systems At the core of the infrared system is the detector, which may either be a single detector or an array detector.

1. *Single Detectors*: The single detector consists of a single element comprising one of the sensors discussed in the preceding section. The image of an object is obtained by using a scanner, which sequentially focuses different sections of the object unto the detector. When it is only necessary to monitor the object along a line, a line scan is used. For area detection, the scan is done both horizontally and vertically. For spot detection, there is no need for scanning, and the detector is focused at one spot.
2. *Linear Detectors*: The linear detector consists of a linear array of detectors, about 4096. Points along a line on the object are thus detected simultaneously, and thus for linear detection, no scanning is necessary. For area detection, the object is scanned vertically, if the array is horizontal.
3. *Array Detectors*: Two-dimensional array detectors are able to acquire an area image of an object without scanning and may have up to 4096×4096 elements.

Many commercial infrared systems use liquid nitrogen to cool the detectors, with the detector being mounted in the base of a Dewar flask. However, more recent developments may involve solid-state detectors that use the detector substrate as a heat sink, thereby making additional cooling unnecessary. During operation, the scanner detects heat from the source by capturing thermal radiation emitted as infrared electromagnetic waves. The use of a chopper within the scanner provides a radiance reference for the system. The detector output level may be set with this reference a number of times per second. This controls system drift that would otherwise be present when temperature difference measurements are made in the presence of a changing background temperature.

The detected infrared energy is converted to an electrical signal and then quantified for each element using either color or different levels of gray to show discrete levels of thermal energy. Image processing may also be performed directly on the electrical signal. Most systems are either of the short wavelength or long wavelength type. The short wavelength systems have detectors that are primarily sensitive to the 1–6 μm range, while the long wavelength detectors are sensitive to the 8–14 μm range. Commercially available long wavelength systems are capable of measuring temperatures in the range of −20 to 2500 °C with a resolution of 0.2 °C.

Two major applications of infrared sensing are currently in use: (i) thermal imaging and (ii) infrared signal analysis.

Thermal Imaging: In thermal imaging, thermal images are recorded on-line, and by image processing, process states such as weld geometry can be measured.

Infrared Signal Analysis: In infrared signal analysis, infrared signals are converted to electrical signals and analyzed using signal processing techniques.

Problem Sources

1. *Plasma Interference*: In addition to radiation from the object to be monitored such as the molten pool or joint, radiation may also be received from the plasma. This would result in erroneous interpretation of the data. For reliable monitoring, the plasma either needs to be screened from the detector or its radiation needs to be filtered from the incident radiation. The use of a detector that is most sensitive in the appropriate wavelength range would increase the signal-to-noise ratio.

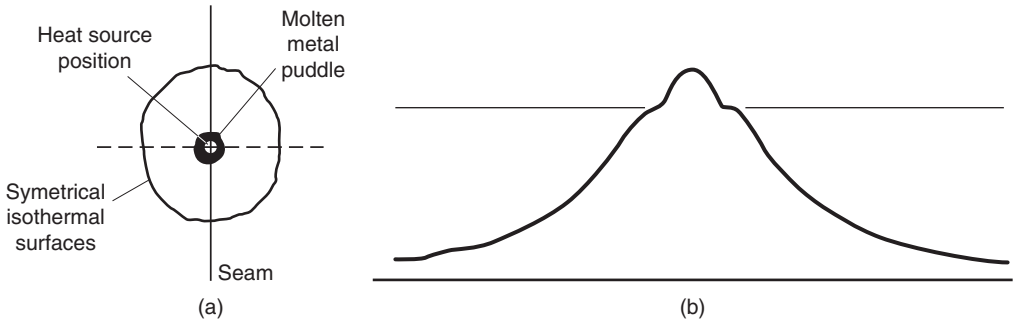


Figure 15.18 (a) Thermal image of a stationary weld pool with heat source directly above joint center. (b) Line scan of the thermal image. Source: From Chin et al. (1983)/with permission of American Welding Society.

2. *Emissivity*: The infrared radiation from a material is a function of both temperature and emissivity. Thus, for a given detected radiation, the absolute temperature cannot be determined unless the emissivity is known. However, the emissivity can vary significantly for a given material, depending on both the surface condition and temperature. For a steel, for example, this may vary from about 0.07 for a polished surface to about 0.79 for an oxidized surface.

Application to Process Monitoring In this section, welding is used as a specific example. However, the concepts discussed are applicable to other processes that involve the formation of a molten pool. A regular welding process, for example, results in symmetric and repeatable patterns of the temperature gradients. Imperfections in the process, however, result in a discernible change in the thermal profiles.

The thermal image of a stationary welding process with the heat source positioned directly above the joint center is illustrated in Figure 15.18a. The corresponding line scan across the center is shown in Figure 15.18b, from which the distribution is found to be symmetrical. The average weld pool diameter can be obtained from the line scan and is given by the inflections around the peak temperature.

When the heat source is shifted to one side of the joint center, the thermal image becomes distorted in shape, consisting then of half-moon shapes (Figure 15.19a,b). This asymmetrical temperature distribution is caused by the excess energy that is deposited on one side of the joint relative to the other and the contact resistance at the joint, which reduces heat flow across the joint, resulting in higher temperatures on the side with excess energy. The heat source could then be moved in the appropriate direction until the two radii are equal.

15.2.4.2 Ultraviolet Detection Radiation associated with the plasma is concentrated in the UV band of the electromagnetic spectrum (0.01–0.4 μm). Thus, since welding plasma is primarily generated from the keyhole, ultraviolet sensing is usually correlated with keyhole formation and weld penetration.

There are two major applications of UV sensing: (i) spectroscopic analysis and (ii) ultraviolet signal analysis.

Ultraviolet Spectroscopic Analysis Spectroscopic analysis is useful for determining the ion content in the plasma. This is important in determining the extent of plasma formation due to metal vapor and shielding gas. Especially important is plasma absorption of the laser power above the workpiece

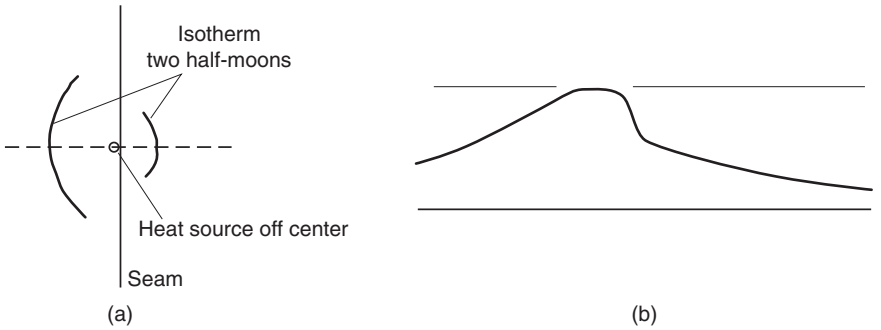


Figure 15.19 (a) Thermal image for a nonsymmetrically located heat source. (b) Corresponding line scan. Source: From Chin et al. (1983)/with permission of American Welding Society.

or coupling of laser energy to the material, a phenomenon that is highly influenced by the plasma size and its spectral characteristics. The spectrum provides information on the plasma electron temperature, T_{pl}, and also the electron density, n_e, which is normally obtained using techniques involving the relative line intensity and stark broadening of spectral lines. These techniques do not need calibration. The spectrum may be obtained using a spectograph coupled with a gated optical multichannel analyzer.

Plasma temperatures can be extracted from spectroscopic information by using relative emission intensity ratio of ion states of an element if the plasma is dense enough for local thermodynamic equilibrium (LTE) to exist.

Ultraviolet Signal Analysis While spectroscopic analysis is very important, recent developments in ultraviolet signal analysis have shown a lot of potential. In ultraviolet signal analysis, ultraviolet signals from the process are converted into electrical signals. Magnitude and frequency analyses of the output signals are then correlated with process states. In laser welding, for example, when penetration occurs, a clear relationship between full penetration and the ultraviolet plasma signal is usually observed.

Fluctuations in the laser-induced plasma do occur even when the laser power is constant. Such fluctuations can be monitored using a photodiode. Sudden changes in mass flow rate (and thus pressure) of the plasma out of a keyhole result in steep and narrow peaks in the photodiode signal. A key application of plasma monitoring is in detection of melt through during laser welding, since this significantly reduces the pressure in the keyhole, and thereby the photodiode signal.

15.2.5 Optical (Vision) Sensing

Optical sensors are often used for monitoring molten pool geometry, kerf size, and observing flow on the free pool surface and chevron formation. In addition, they are also used for seam tracking. The basic components of an optical sensing system include the sensor or detector, illumination source, object, transmission elements, and the processor.

15.2.5.1 Optical Detectors The most common type of optical detector is the solid-state array camera, which usually has an array of charge injection or CCD light-sensitive elements. The output from the camera is dumped into a memory buffer for analysis. An optical filter is usually used to separate the laser radiation from heat radiation. The detector elements may require significant attenuation, especially when used for a focused beam.

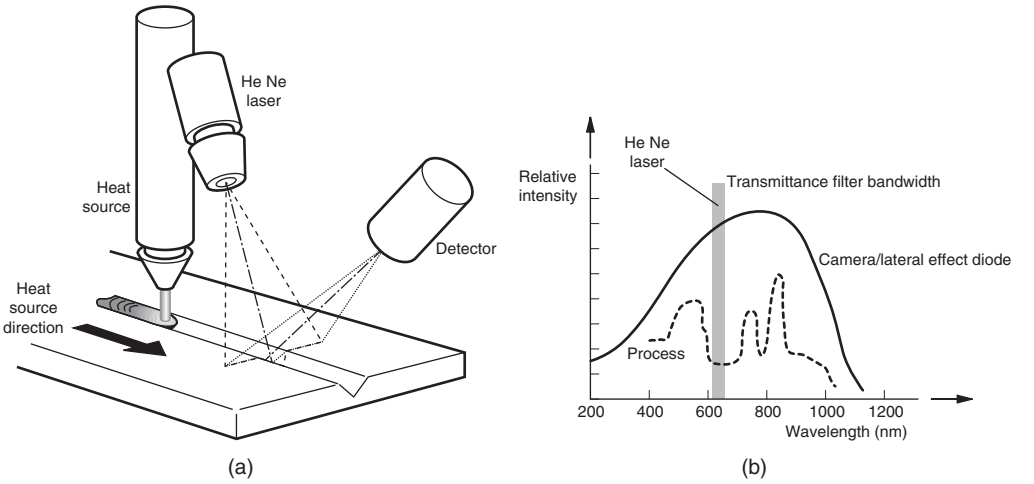


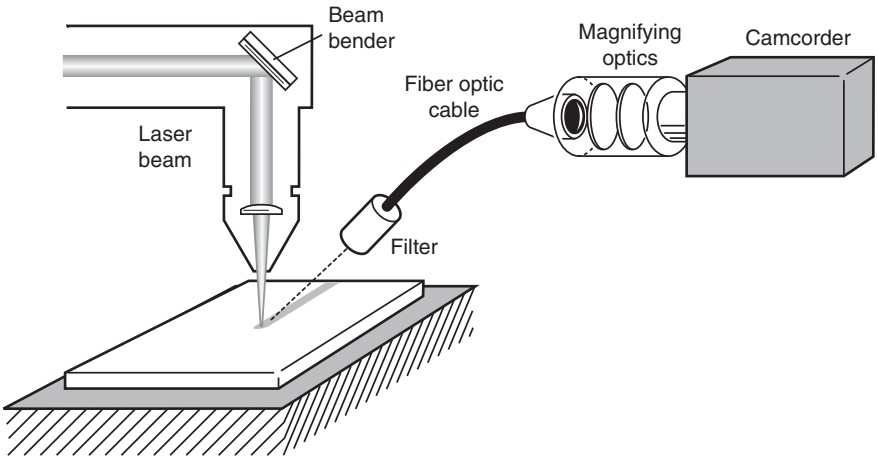
Figure 15.20 (a) Illustration of a structured light set-up. (b) Illustration of detector and process spectral characteristics. Source: From Dornfeld (1983)/The American Society of Mechanical Engineers.

15.2.5.2 Detector Setup Some of the problems associated with optical sensing include the extreme brightness of the plasma plume compared to that of the molten pool (high contrast) and dependence of the intensity on processing conditions. Spatter and fumes may also obscure the object to some extent. Separate illumination is thus normally used to enhance the system resolution. It may be in the form of either structured light or general (nonstructured) illumination. A structured light is a pattern of lines or a grid of light that is projected unto the object to help provide information on the three-dimensional shape of the object, based on the apparent distortion of the pattern (Figure 15.20a).

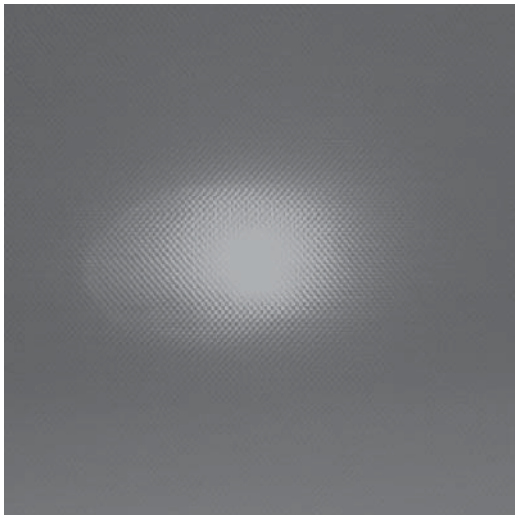
The general illumination could come from an auxiliary high-intensity light source. An example involves illuminating the object with a narrow bandwidth laser beam, with the beam bandwidth selected to be in the region where the sensitivity of the detector is high, based on the spectral characteristics of the detector (Figure 15.20b). This is because the light spectrum from an object generally does not have equal intensity at all wavelengths (Figure 15.16). All light incident on the detector is then filtered except for the narrow bandwidth of the auxilary beam, thereby subduing the effect of the bright light from the plume.

The detector arrangement for an optical sensing system is similar to that of an infrared array (Section Infrared Systems). It may be a linear or two-dimensional array.

15.2.5.3 Edge Detection Methodology Edge detection is used to identify the molten pool boundary. This enables say the weld bead size to be determined, and hence, an estimate to be made of the load-carrying capacity of the weldment. Using the setup shown in Figure 15.21a, an example molten pool image data as obtained from the camera are shown in Figure 15.21b. Figure 15.22 shows a three-dimensional view of the same image. The image could be contaminated by noise from a number of sources.

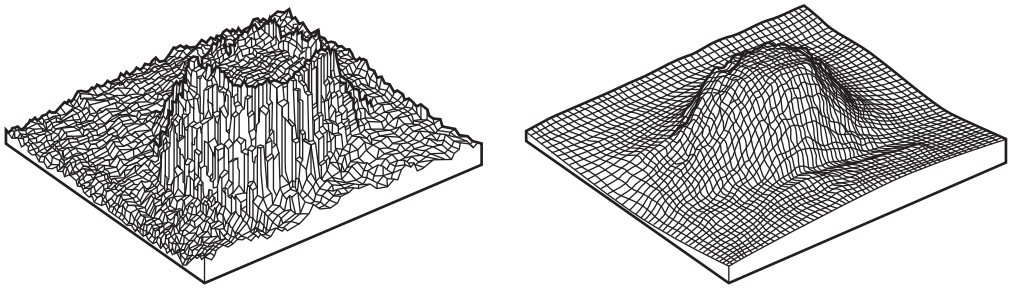


(a)



(b)

Figure 15.21 (a) Vision monitoring system. (b) Typical weld pool image before image processing. Source: From Tsai (1998)/University of Michigan.



(a) (b)

Figure 15.22 (a) Three-dimensional view of the weld pool image in Figure 15.21. (b) Three-dimensional view of weld pool after nine-point averaging. Source: From Tsai (1998).

Noise Sources Potential sources of noise include the following:

1. *Dust on the Optical Components*: This results from condensation of metal vapor.
2. *Noise Due to Fiber Thread Boundary*: When a fiber optic cable is used, the image at the end of the cable is a combination of small images transmitted by individual fibers in the bundle. Very thin lines encircle each individual image cell, representing the boundaries of each fiber. These influence the image quality because they result in uneven brightness, which is a noise source on the image, and cannot be eliminated.
3. *Noise from the CCD Array*: Each individual photo diode cell of the CCD array converts a portion of the image into an electrical signal. Increasing the density of photo diodes in one CCD array results in a more vivid image. The image resolution is determined by the number of photo diodes used. The noise results when the CCD array converts the two-dimensional image into electrical signals. This is a minor noise source that cannot be corrected.
4. *Image Digitization Noise*: When the resolution of the frame grabber (say, 640×480 pixels) is different from that of the camcorder (say, 512×512 pixels) that can be a source of error.

As a result of these error sources, low-pass filtering is necessary.

Filtering There are different types of filters available for analyzing optical signals. As an example, the nine-point averaging method, which is commonly used, is discussed. It involves replacing a given point $P(i, j)$ by the average of all eight neighboring points, plus the original value of the point:

$$P(i,j)_{\mathrm{n}} = \frac{1}{9}\left[\sum_{j-1}^{j+1}\sum_{i-1}^{i+1} P(i,j)\right] \tag{15.13}$$

where $P(i, j)_{\mathrm{n}}$ is the new value of the point $P(i, j)$. Application of nine-point averaging to the original image of Figure 15.22a is shown in Figure 15.22b.

Edge Detection Following filtering, edge detection is then used to determine, for example, the molten pool boundaries. Two common techniques are directional and nondirectional edge detection.

Directional Edge Detection: The gradient value of a point $P(i, j)$ in the x-direction for directional edge detection is given by

$$\begin{aligned} x \text{ gradient of } P(i,j) = |&P(i-1,j-1) + P(i-1,j) + P(i-1,j+1) - P(i+1,j-1) \\ &- P(i+1,j) - P(i+1,j+1)| \end{aligned} \tag{15.14a}$$

Likewise, the gradient value in the y-direction is

$$\begin{aligned} y \text{ gradient of } P(i,j) = |&P(i-1,j-1) + P(i,j-1) + P(i+1,j-1) - P(i-1,j+1) \\ &- P(i,j+1) - P(i+1,j+1)| \end{aligned} \tag{15.14b}$$

The absolute value of the gradient is used, to accommodate both positive and negative values. Boundary points are determined using a threshold, and the following equation may be used to identify candidate boundary points:

This is a candidate boundary point, if

(x-directional gradient $\geq$ threshold) or (y-directional gradient $\geq$ threshold) (15.15)

Nondirectional Edge Detection: The nondirectional edge detector is based on the gradient of an image $f(x, y)$ at location $P(x, y)$:

$$\nabla f = \begin{pmatrix} \nabla_x f \\ \nabla_y f \end{pmatrix} = \begin{pmatrix} \dfrac{\partial f}{\partial x} \\ \dfrac{\partial f}{\partial y} \end{pmatrix} \tag{15.16}$$

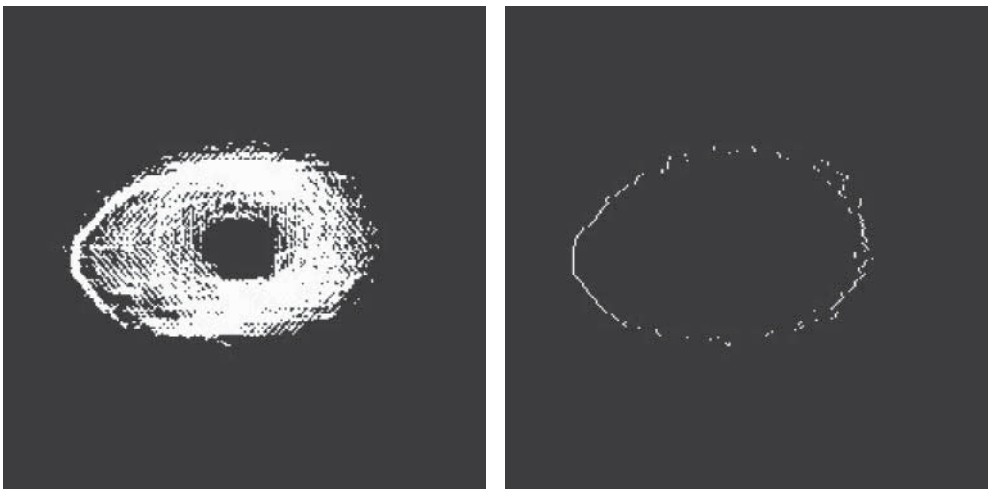
where $\nabla_x f = \frac{\partial f}{\partial x}$ is the x-directional gradient and $\nabla_y f = \frac{\partial f}{\partial y}$ is the y-directional gradient.

The gradient vector points in the direction of maximum change rate. The x- and y-directional gradient operators can be calculated using Sobel operators, which provide smoothing during the differentiation operation. Sobel operators in the x and y directions are represented as

$$\nabla_x f = |P(i-1,j-1) + 2P(i-1,j) + P(i-1,j+1) - P(i+1,j-1) - 2P(i+1,j) - P(i+1,j+1)| \tag{15.17a}$$

$$\nabla_y f = |P(i-1,j-1) + 2P(i,j-1) + P(i+1,j-1) - P(i-1,j+1) - 2P(i,j+1) - P(i+1,j+1)| \tag{15.17b}$$

The values are always positive. Figure 15.23a shows the results (in two dimensions) of applying the nondirectional gradient detector to the image in Figure 15.22b.



(a) (b)

Figure 15.23 (a) Results of application of nondirectional edge detector. (b) Weld pool boundary points determined using the line-by-line searching method. Source: From Tsai (1998)/University of Michigan.

Finding the Boundary Points The actual boundary points need to be determined from the candidate points of Figure 15.23a. A simple method is the line-by-line searching algorithm which searches the entire image row-by-row. One criterion for selecting outer boundary points might be

outer boundary point

= outer candidate point which has at least two other candidate points in its neighborhood (15.18)

The resulting boundary for the image in Figure 15.23a is shown in Figure 15.23b.

Calibration Calibration is used to convert the initial unit of pixel to metric units and to determine the threshold values used in the boundary point search algorithm. This may be done experimentally by first placing white circles of known diameter against a black background and capturing their images. The resulting images may be elliptical, depending on the camera orientation. The lengths of the major and minor axes of the image are then determined in pixels. From these measurements, the image resolution can be determined in millimeters per pixel.

15.3 SUMMARY

Sensors for monitoring laser processes can be categorized into three groups:

Beam monitoring sensors: They are used for measuring beam characteristics such as power and beam distribution. Examples of such sensors are the beam dump (for power) and laser beam analyzers (rotating rod, hollow needle, or camera) for beam distribution. The beam size can also be determined using a Kapton film.

Process monitoring sensors: These are used to measure actual process features such as geometry of the molten pool, kerf size, and defects that are formed during the process. The top view of the pool geometry, for example, can be measured using vision, infrared, or ultraviolet sensors. Infrared and ultraviolet detectors can also provide information on the stability and therefore quality of the process. The individual detector elements can be arranged either in a linear array (for measurement along a line) or a two-dimensional array (for area images). Whereas the vision, infrared, and ultraviolet sensors can detect defects that are exposed to the surface, AE detectors can, in addition, detect defects that are embedded in the workpiece. In all cases, special precautions have to be taken to prevent contamination of the measured signal by noise.

Seam tracking sensors: These are used to detect the joint along which the laser beam is expected to follow, enabling the beam to be guided automatically. Seam tracking can be done using vision and infrared sensors.

APPENDIX 15.A

List of symbols used in the chapter

Symbol	Parameter	Units
c_d	Damping coefficient	N s/m
c_1	Constant $= 2\pi c^2 h_p = 3.74 \times 10^{-16}$	J m^2/s
c_2	Constant $= ch_p/k_B = 1.44 \times 10^{-2}$	m K

Symbol	Parameter	Units
d_h	Diameter of hole in Kapton film	mm
ΔH_v	Energy necessary to heat and vaporize a unit volume of material	J/mm^3
I	Total hemispherical radiation intensity	W/m^2
I_λ	Amplitude of radiant energy	J/m^3 s
k_s	Spring constant	N/m
$\dot{m}$	Mass flow rate of fluid	kg/s
ND	Neutral density number	—
$P(i, j)$	Location of a point representing a pixel	—
$P(i, j)_n$	New value of the point $P(i, j)$	—
Q_e	Total pulse energy	J
RMS	Root mean square value of signal	V
T	Transmission ratio	—
T_{pl}	Plasma temperature	K
$V(t)$	Signal voltage function	V
w_b	Base radius of laser beam intensity distribution	μm
w_t	Top radius of laser beam intensity distribution, assumed to be truncated	μm
x	Displacement of mass	mm
y	Input displacement of base frame	mm
Y	Maximum amplitude of the input base displacement	mm
z	Relative displacement between base and mass	mm
Z	Maximum amplitude of the relative motion between the base and mass	mm
α_A	Absolute Seebeck coefficient of material A	V/K
α_B	Absolute Seebeck coefficient of material B	V/K
ϕ	Phase angle between input motion y and relative motion z	rad
λ_m	Wavelength at which peak intensity occurs	μm
ω	Angular frequency if input base oscillation	rad/s
ω_n	Natural frequency of vibration of undamped system	rad/s
ω_v	Angular velocity	rad/s
ζ	Damping ratio	—

PROBLEMS

15.1. A laser beam of power 3 kW is dumped into a cone calorimeter beam dump. If the inlet temperature of the water flowing through the calorimeter is 30 °C, determine the water flow rate through it if the outlet temperature cannot exceed 95 °C. The specific heat of water is 4.186 J/g °C = 4186 J/kg °C.

15.2. Determine the amplitude of radiant energy at a wavelength of 2 μm by a blackbody at a temperature of 1500 °C.

15.3. The ratio of output amplitude Z (measured by the sensor) to input amplitude Y (from the induced vibrations) for an AE piezoelectric transducer is given by

$$\frac{Z}{Y} = \frac{\omega 2}{{\omega_n}^2}\left[\left(1 - \frac{\omega^2}{{\omega_n}^2}\right)^2 + \left(2\zeta\frac{\omega}{\omega_n}\right)^2\right]^{-\frac{1}{2}}$$

(a) Show that the above equation can be transformed to

$$\frac{Z}{Y} = \frac{m\omega^2}{\sqrt{(k_s - m\omega^2)^2 + (c_d\omega)^2}}$$

with the natural frequency given by

$$\omega_n = \sqrt{\frac{k_s}{m}}$$

and the damping ratio given by

$$\zeta = \frac{c_d}{2m\omega_n}$$

(b) A new adhesive is developed that, if used to couple the piezoelectric sensor to the object's surface, suppresses all damping giving a damping ratio $\zeta \approx 0$. What are the effects for process monitoring? Should this adhesive be used? (Hint: think about the ratio of amplitudes in case of $\omega = \omega_n$ and the phase angle φ).

BIBLIOGRAPHY

Astic, D., Canosa, A., Vigiano, P., and Autric, M. (1991). Visible spectroscopy of laser produced aluminum plasma. Electron temperature and density determinations. *ICALEO* 1991: 94–103.

Autric, M., Vigiano, P., and Astic, D. (1988). Visible spectroscopy of laser produced plasma. *SPIE High Power CO2 Laser Systems and Applications* 1020: 103–112.

Benedict, R.P. (1984). *Fundamentals of Temperature, Pressure, and Flow Measurements*, 3e. New York: Wiley.

Beyer, E., Abels, P., Drenker, A. et al. (1991). New devices for on line process diagnostics during laser machining. *ICALEO* 1991: 133–139.

Carlson, N.M. and Johnson, J.A. (1988). Ultrasonic sensing of weld pool penetration. *Welding Journal* 67: 239s–246s.

Chen, H.B., Li, L., Brookfield, D.J. et al. (1991). Laser process monitoring with dual wavelength optical sensors. *ICALEO* 1991: 113–122.

Chen, S.L., Li, L., Modern, P., and Steen, W.M. (1991). In process laser beam position sensing. *SPIE Industrial and Scientific Uses of High-Power Lasers* 1502: 123–134.

Chin, B.A., Madsen, N.H., and Goodling, J.S. (1983). Infrared thermography for sensing the arc welding process. *Welding Journal* 62: 227s–234s.

Denney, P.E. and Metzbower, E.A. (1991). Synchronized laser-video camera system study of high power laser material interactions. *ICALEO* 1991: 84–93.

DiPietro, P. and Yao, Y.L. (1994). An investigation into characterizing and optimizing laser cutting quality – A review. *International Journal of Machine Tools & Manufacture* 34 (3): 225–243.

Dornfeld, D.A. (1983). Arc weld monitoring for process control – Seam tracking. In: *Computer-Integrated Manufacturing*, ASME Bound Volume (ed. M. Martinez), 89–97. The American Society of Mechanical Engineers.

Dove, R.C. and Adams, P.H. (1964). *Experimental Stress Analysis and Motion Measurement*. Columbus, OH: Charles E. Merrill Publishing Co.

Farson, D.F., Fang, K.S., and Kern, J. (1991). Intelligent laser welding control. *ICALEO* 1991: 104–112.

Gatzweiler, W., Maischner, D., and Beyer, E. (1988). On-line plasma diagnostics for process control in welding with CO_2 lasers. *SPIE High Power CO2 Laser Systems and Applications* 1020: 142–148.

Gregersen, O. and Olsen, F.O. (1990). Beam analyzing system for CO_2 lasers. *ICALEO* 28–35.

Hamann, C., Rosen, H.-G., and LaBiger, B. (1989). Acoustic emission and its application to laser spot welding. *SPIE High Power Lasers and Laser Machining Technology* 1132: 275–281.

Hettwer, M. (2011). Laser beam profiling in the production process. *Laser Technik Journal* 1: 34–37.

Holtgen, B., Treusch, H.G., and Aehling, H. (1988). Diagnostic system for lasers in the visible and near infrared region. *SPIE Laser Assisted Processing* 1022: 52–54.

Hossain, M.A., Canning, J., Cook, K., and Jamalipour, A. (2015). Smartphone laser beam spatial profiler. *Optics Letters* 40 (22): 5156–5159.

Kannatey-Asibu, E. Jr. and Dornfeld, D.A. (1981). Quantitative relationships for acoustic emission from orthogonal metal cutting. *ASME Journal of Engineering for Industry* 103: 330–340.

Kannatey-Asibu, E. Jr. (1982). On the application of the pattern recognition method to manufacturing process monitoring. In: *Proceedings of the 10th North American Manufacturing Research Conference*, 487–492.

Lewis, G.K. and Dixon, R.D. (1985). Plasma monitoring of laser beam welds. *Welding Journal* 64: 49s–54s.

Li, L., Qi, N., Brookfield, D.J., and Steen, W.M. (1990). On-line laser weld sensing for quality control. *ICALEO* 1990: 411–421.

Li, L. and Steen, W.M. (1992). Non-contact acoustic emission monitoring during laser processing. *ICALEO* 1992: 719–728.

Lim, G.C. and Steen, W.M. (1982). Measurement of the temporal and spatial power distribution of a high-power CO_2 laser beam. *Optics and Laser Technology* 14 (3): 149–153.

Lim, G.C. and Steen, W.M. (1984). Instrument for instantaneous in situ analysis of the mode structure of a high-power laser beam. *Journal of Physics E: Scientific Instruments* 17: 999–1007.

Liu, X. and Kannatey-Asibu, E. Jr. (1990). Classification of AE signals for martensite formation from welding. *Welding Journal* 69 (10): 389s–394s.

Liu, Y. and Leong, K.H. (1992). Laser beam diagnostics for kilowatt power pulsed YAG laser. *ICALEO* 1992: 77–87.

Maischner, D., Drenker, A., Seidel, B. et al. (1991). Process control during laser beam welding. *ICALEO* 1991: 150–155.

Maxfield, B.W., Kuramoto, A., and Hulbert, J.K. (1987). Evaluating EMAT designs for selected applications. *Materials Evaluation* 45: 1166–1183.

Mombo-Caristan, J.-C., Koch, M., and Prange, W. (1993). Process controls for laser blank welding. *ALAW* 1–7.

Orlick, H., Morgenstern, H., and Meyendorf, N. (1991). Process monitoring in welding and solid state lasers by sound emission analysis. *Welding and Cutting* 12: 15–18.

Parthasarathi, S., Khan, P.A.A., and Paul, A.J. (1992). Intelligent laser processing of materials. *ICALEO* 1992: 708–718.

Plass, W., Maestle, R., Wittig, K. et al. (1996). High-resolution knife-edge laser beam profiling. *Optics Communications* 134: 21–24.

Richardson, R.W., Gutow, D.A., Anderson, R.A., and Farson, D.F. (1984). Coaxial arc weld pool viewing for process monitoring and control. *Welding Journal* 63: 43s–50s.

Sasnett, M.W. (1990). LIA beam characterization report. *ICALEO* 1990: 1–12.

Sheikh, M. and Riza, N.A. (2009). Demonstration of pinhole laser beam profiling using a digital micromirror device. *IEEE Photonics Technology Letters* 21 (10): 666–668.

Shershulin, V.A., Samoylenko, S.R., Sedov, V.S. et al. (2017). Using Si-doped diamond plate of sandwich type for spatial profiling of laser beam. *Laser Physics Letters* 14, 026003 (4pp.).

Smith, E.T. (1999). Monitoring laser weld quality using acoustic signals. PhD Dissertation. University of Michigan, Ann Arbor, Michigan.

Sokolowski, W., Herziger, G., and Beyer, E. (1989). Spectroscopic study of laser-induced plasma in the welding process of steel and aluminum. *SPIE High Power Lasers and Laser Machining Technology* 1132: 288–295.

Steen, W.M. and Weerasinghe, V.M. (1986). In-process beam monitoring. *SPIE Laser Processing: Fundamentals, Applications, and Systems Engineering* 668: 37–44.

Steen, W.M. and Mazumder, J. (2010). *Laser Material Processing*. Berlin: Springer-Verlag.

Sun, A., Kannatey-Asibu, E. Jr., and Gartner, M. (1999). Sensor systems for real time monitoring of (laser) weld quality. *Journal of Laser Applications* 11 (4): 153–168.

Sun, A., Kannatey-Asibu, E. Jr., and Gartner, M. (2002). Monitoring of laser weld penetration using sensor fusion. *Journal of Laser Applications* 14 (2): 114–121.

Tsai, F.-R. (1998). Vision sensing, modeling, and control of laser weld pool geometry. PhD Dissertation. University of Michigan, Ann Arbor, Michigan.

VanderWert, T.L. (1992). In-process workpiece and fixture sensing for flexible, repeatable laser processing of 3D parts. *ICALEO* 1992: 699–707.

Voelkel, D.D. and Mazumder, J. (1990). Visualization and dimensional measurement of the laser weld pool. *ICALEO* 1990: 422–429.

Wang, Z.Y., Liu, J.T., Hirak, D.M. et al. (1991). Measurement of pulsed laser beam dimensions using KAPTON films. *ICALEO* 1991: 74–83.

Weeter, L. and Albright, C. (1987). The effect of full penetration on laser-induced stress wave emissions during laser spot welding. *Materials Evaluation* 45: 353–357.

Whitehouse, D.R. and Nilsen, C.J. (1990). Plastic burn analysis (PBA) for CO_2 laser beam diagnostics. *ICALEO* 1990: 13–27.

16 Processing of Sensor Outputs

In Chapter 15, the major sensors that are used in monitoring laser processes were discussed. Process outputs, as reflected in the measured signals, are generally stochastic in nature. Thus, it is often necessary to further analyze the signals to extract pertinent information about the state of the process. There are several methods available for analyzing signals, but only a basic approach is considered here. The discussion in this chapter on processing of sensor outputs therefore focuses on the identification of faults that may occur during laser processing, for example, cracking, porosity, no weld, and so on, and is limited to classification methods. Implementation of a classification scheme normally involves three steps (Figure 16.1):

1. Sampling (digitizing) the input signal to produce the pattern space.
2. Feature extraction, which normally involves transformation of the signal from the pattern into the feature space, from which useful information can be obtained. This is usually accompanied by a reduction in the data size.
3. Classification of the feature space to identify the individual signal sources (classes).

Since sampling of an analog signal influences the results of its transformation, the discussion on these two issues are combined into one, under signal transformation. The Fourier transform, the discrete Fourier transform, and their properties are discussed, along with pitfalls of digital analysis. This is followed by the sampling theorem, aliasing, and leakage. Next, data reduction is presented, focusing on the variance and Fisher criteria. Pattern recognition and neural network analysis are then discussed in terms of pattern classification. Finally, time–frequency methods such as the short-time Fourier transform (STFT) are briefly outlined.

16.1 SIGNAL TRANSFORMATION

Measured signals are often contaminated by noise. To characterize the system from which the signal is being obtained, in the presence of the noise, the signal is first transformed into another domain to facilitate analysis and interpretation. There are various transforms that can be used, such as Fourier transform, the Walsh–Hadamard transform, the Haar transform, and the Karhunen–Loeve transform. The Fourier transform is perhaps the most commonly used one because it permits a more useful engineering interpretation of the results and will therefore be the focus of our discussion.

16.1.1 The Fourier Transform

The Fourier transform involves the transformation of information from the time domain (time series) to the frequency domain (spectrum). The basis of spectral analysis is that a nonperiodic signal, $f(t)$,

Principles of Laser Materials Processing: Developments and Applications, Second Edition. Elijah Kannatey-Asibu, Jr.
© 2023 John Wiley & Sons, Inc. Published 2023 by John Wiley & Sons, Inc.
Companion Website: www.wiley.com/go/Kannatey-Asibu/PrinicplesLaserMaterialsProcessing

Figure 16.1 Basic classification scheme.

that is bounded and has a finite number of discontinuities can be represented in the frequency domain using the Fourier transform, which is given as:

$$g(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt \tag{16.1}$$

and

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} g(\omega)e^{i\omega t}d\omega \tag{16.2}$$

where $i = \sqrt{-1}$, $\omega = 2\pi\nu$ is the angular frequency and ν is the frequency.

Equation (16.1) is the Fourier transform of $f(t)$ and Eq. (16.2) is the inverse Fourier transform, and together, they constitute the Fourier transform pair. The Fourier transform is a special form of the Fourier series and is applicable to nonperiodic functions. Note that there are several variants of the Fourier transform pair.

$g(\omega)$ is a complex continuous function of ω and can be rewritten as:

$$g(\omega) = A(\omega) + iB(\omega) \tag{16.3}$$

where $A(\omega)$ and $B(\omega)$ are the real and complex components, respectively, of $g(\omega)$.

The amplitude spectrum, $a(\omega)$, of the function $f(t)$ is then given by:

$$a(\omega) = |g(\omega)| = \sqrt{[A(\omega)]^2 + [B(\omega)]^2} \tag{16.4a}$$

the power spectrum by:

$$q(\omega) = |g(\omega)|^2 = [A(\omega)]^2 + [B(\omega)]^2 \tag{16.4b}$$

and the phase spectrum by:

$$\varphi(\omega) = \tan^{-1}\frac{B(\omega)}{A(\omega)} \tag{16.4c}$$

Direct use of the amplitude spectrum, $a(\omega)$, can sometimes present problems when there are pseudorandom components that can produce varying $a(\omega)$ that average out to zero over a long signal length, $\Delta\tau$. The use of the power spectrum eliminates this problem. To make the power spectrum independent of the duration, $\Delta\tau$, of the data, it can be reexpressed as:

$$q_n(\omega) = \frac{1}{\Delta\tau}|g(\omega)|^2 \tag{16.5}$$

16.1.2 The Discrete Fourier Transform (DFT)

For digital analysis, the discrete equivalent of the Fourier transform is used. Let $\{x(j)\}$ denote a sequence $x(j), j = 0, 1, \ldots, N_{m-1}$ of N_m finite valued real or complex numbers, representing a time

series of data. The discrete Fourier transform of the sequence is given as:

$$y(k) = \frac{1}{N_m} \sum_{j=0}^{N_m-1} x(j) e^{\frac{-i2\pi jk}{N_m}} \qquad k = 0, 1, \ldots, N_m - 1 \tag{16.6}$$

and the inverse transform:

$$x(j) = \sum_{k=0}^{N_m-1} y(k) e^{\frac{i2\pi jk}{N_m}} \qquad j = 0, 1, \ldots, N_m - 1 \tag{16.7}$$

To illustrate some basic concepts of the DFT, a real-valued signal and its DFT are shown in Figure 16.2. In this figure,

$$\nu_0 = \frac{1}{\Delta\tau}$$

$$\Delta\tau = N_m \delta\tau$$

$$\nu_s \text{ (sampling rate)} = N_m \nu_0 = \frac{1}{\delta\tau}$$

$$\nu_f \text{ (folding frequency)} = \frac{\nu_s}{2}$$

The signal time series, $x(j)$, is considered to be periodic. Some points to note in relation to the DFT are as follows:

1. Each component of the transform is a harmonic. The actual frequency is $k\nu_0$, where ν_0 is the fundamental frequency.
2. For real $x(j)$, the real part of the digital transform, Re$[y(k)]$, is symmetric about the folding frequency $\left(\nu_f = \frac{\nu_s}{2}\right)$ where ν_s is the sampling rate. The imaginary component, Im$[y(k)]$, is antisymmetric.

16.1.3 Pitfalls of Digital Analysis

To understand some of the problems one might encounter in using the discrete Fourier transform, and consider a simple cosine waveform. The continuous transform of this waveform consists of two impulse functions which are symmetric about zero frequency (Figure 16.3a). For digital analysis, only a finite portion of the original cosine waveform signal is observed. This is equivalent

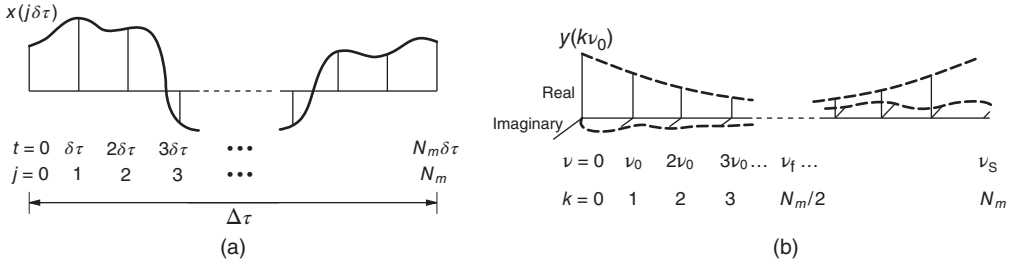


Figure 16.2 (a) A real signal. (b) Its complex discrete Fourier transform. Source: Bergland (1969)/Interdisciplinary Centre for Mathematical and Computational Modelling.

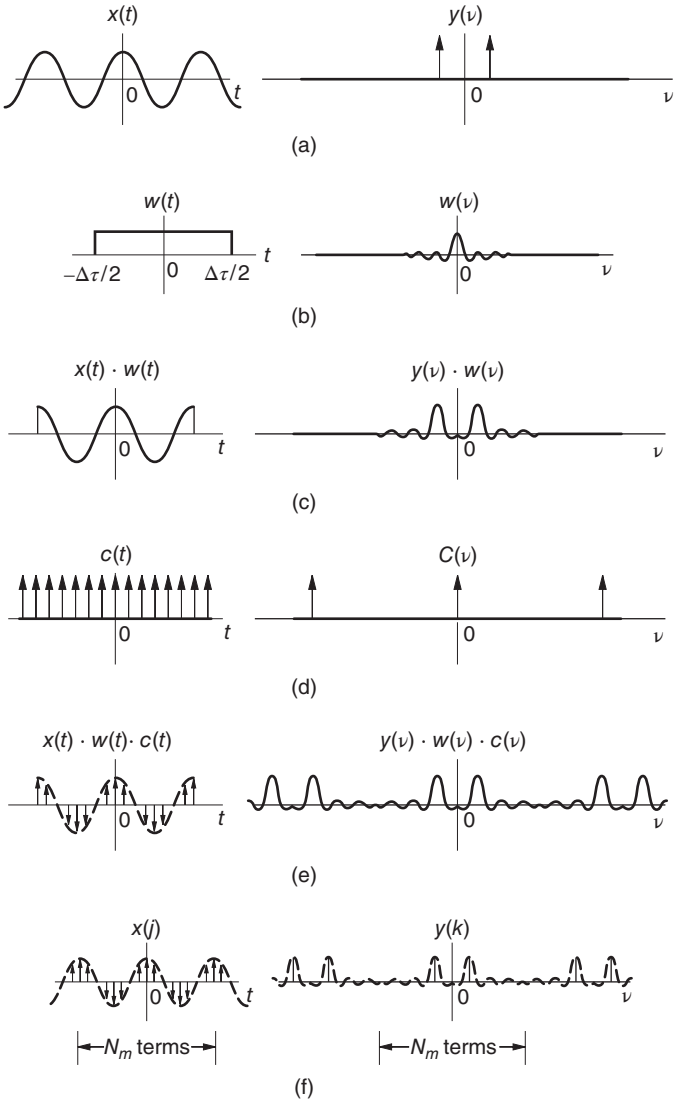


Figure 16.3 The Fourier transform of a finite length continuous cosine waveform. (a) Original analog signal and its transform. (b) Window function to be applied. (c) Windowed signal. (d) Sampling points. (e) Digitized or sampled windowed signal. (f) Sampled signal and its continuous frequency domain function. Source: Bergland (1969)/Interdisciplinary Centre for Mathematical and Computational Modelling.

to multiplying the signal by a unity rectangular window. The transform of this rectangular window is a (sin x/x) function (Figure 16.3b). This results in distortion of the transform of the finite waveform (Figure 16.3c). Now sampling is equivalent to multiplying the signal by a train of impulses (Figure 16.3d). The final transform of the finite sampled waveform is thus a repetition of the distorted transform in the frequency domain (Figure 16.3e,f). What this means is that a simple and straightforward sampling of a signal may introduce spurious frequency components where none existed.

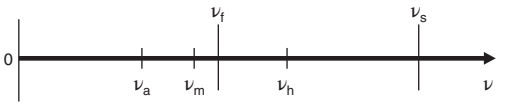


Figure 16.4 Illustration of aliasing on a frequency scale.

16.1.4 The Sampling Theorem

The sampling theorem determines the lower limit of the sampling rate that needs to be used to avoid problems that can result from sampling. For the original continuous signal to be recoverable from its digital version, the digital form has to be obtained by sampling at a frequency which is at least twice the highest frequency in the original continuous signal, that is,

$$\nu_s = 2\nu_h \tag{16.8}$$

where ν_h is the highest frequency in the signal, $\nu_s = 1/\delta\tau$ is the sampling rate (sampling frequency), and $\delta\tau$ is the sampling period.

The sampling theorem only establishes an ideal limit. In practice, the sampling rate has to be about 5–10 times the highest frequency. Some consequences of sampling, or more specifically, of using the discrete Fourier transform are aliasing and leakage. These are briefly discussed in Sections 16.1.5 and 16.1.6.

16.1.5 Aliasing

Aliasing is a phenomenon that happens when a signal is sampled at a rate that is lower than twice the highest frequency (ν_h in Figure 16.4). When this occurs, frequencies above the folding frequency $\nu_f = \frac{\nu_s}{2}$ are aliased or folded back into the frequency range below the folding frequency. In other words, a frequency component that is, say, 100 Hz higher than the folding frequency will be impersonated by a frequency component that is 100 Hz lower than the folding frequency. The low-frequency component has the same amplitude as the high-frequency component. In Figure 16.5, the original high-frequency signal is impersonated by the low-frequency alias which also fits the sampled data. In general,

$$\nu_{alias} = \nu_{sample} - \nu_{actual} \tag{16.9a}$$

or as illustrated in Figure 16.4,

$$\nu_h - \nu_f = \nu_f - \nu_a \tag{16.9b}$$

where ν_a is the aliased frequency and ν_f is the folding frequency.

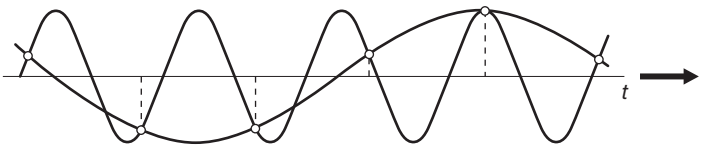


Figure 16.5 An example of low-frequency "impersonating" a high frequency. Source: Bergland (1969)/Interdisciplinary Centre for Mathematical and Computational Modelling.

In other words, the aliasing signal frequency is always the difference between the actual signal and a harmonic (multiple) of the sampling frequency. In Figure 16.4, ν_m is sampled without any problems, since it is lower than half the sampling rate.

16.1.6 Leakage

Leakage results primarily from the fact that the signal being analyzed is finite in length. As indicated in Section 16.1.5, the finite length of data obtained by sampling over a period, $\Delta\tau$, is equivalent to multiplying the signal by a rectangular window. This is the same as a convolution in the frequency domain. Since the rectangular window has a sin x/x function, the convolved signal is also of this form and centered at the spectrum of the original signal. The sin x/x function has a series of spurious peaks that decay very slowly. These spurious peaks (side lobes) introduce high-frequency components into the original signal and those may create aliasing problems where there was none.

The effect of the leakage problem can be minimized by applying a window which has lower side lobes and which is as close as possible to an impulse function. Ideally, this would require an infinite length of data. In lieu of this, windows which have the following characteristics may be used:

1. Main lobe of the window is as narrow as possible.
2. Maximum side lobe is relatively small.

16.2 DATA REDUCTION

A prime consideration in the development of any real-time monitoring system is the response speed, since it determines how soon corrective action can be taken. The response speed in turn depends on the computational efficiency (how fast the appropriate calculations can be made). To improve the computational efficiency, the data size or dimensionality may be reduced by eliminating the feature components that convey the least information about the system. This stage of analysis is referred to as data reduction or feature selection. Data reduction is used to reduce the initial size of the data to make subsequent computations faster, with minimal loss of information.

Here, we consider two criteria for data reduction:

1. Variance criterion.
2. Fisher criterion.

16.2.1 Variance Criterion

The variance criterion involves eliminating the data components with the smallest eigenvalues of the covariance matrix of the sampled data set, or the feature components with the smallest variances.

16.2.2 Fisher Criterion

The Fisher criterion enables the data size to be reduced by selecting the features that minimize scatter of data within the individual classes, while at the same time increasing the scatter (or separation) between the classes, as shown in Figure 16.6. It involves first determining the feature mean for each class, $\bar{\mathbf{y}}_\mathbf{i}$:

$$\bar{\mathbf{y}}_\mathbf{i} = \frac{1}{M_i} \sum_{k=1}^{M_i} \mathbf{y}_\mathbf{ik} \qquad i = 1, \ldots, s \tag{16.10}$$

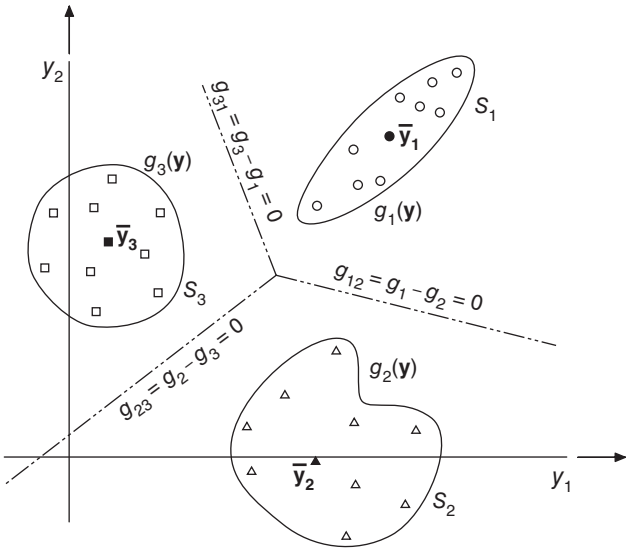


Figure 16.6 A two-dimensional, three-class feature space with linear decision boundaries.

where M_i is the number of patterns in class s_i. In this section and Section 16.3.1, the subscript i is used to indicate the class of the system. $\mathbf{y_{ik}} \equiv \mathbf{y}$ is the kth pattern in class s_i.

The classes of interest in this case would normally include the signal sources, for example, good weld, cracking, porosity, no weld, and any other defects of interest. All other unwanted signals can then be grouped together and classified as noise.

The overall system mean, $\bar{\mathbf{y}}$, is then determined as:

$$\bar{\mathbf{y}} = \sum_{i=1}^{s} p_i \bar{\mathbf{y}}_\mathbf{i} \tag{16.11}$$

where p_i is the *a priori* probability of class s_i.

The scatter within each class is obtained by calculating the covariance matrix $\mathbf{\Phi_{yi}}$ as:

$$\mathbf{\Phi_{yi}} = \frac{1}{M_i} \sum_{k=1}^{M_i} (\mathbf{y_{ik}} - \bar{\mathbf{y}}_\mathbf{i})(\mathbf{y_{ik}} - \bar{\mathbf{y}}_\mathbf{i})' \tag{16.12}$$

which leads to an overall system covariance matrix $\mathbf{\Phi_y}$ of

$$\mathbf{\Phi_y} = \sum_{i=1}^{s} p_i \mathbf{\Phi_{yi}} \tag{16.13}$$

The scatter $\mathbf{\Phi_{ys}}$ between the individual classes is defined as:

$$\mathbf{\Phi_{ys}} = \sum_{i=1}^{s} p_i (\overline{y_i} - \overline{y})(\overline{y_i} - \overline{y})' \tag{16.14}$$

and from this equation the feature selection criterion is defined as:

$$\mathbf{Q} = \frac{\mathbf{\Phi_{ys}}(j,j)}{\mathbf{\Phi_y}(j,j)} \tag{16.15}$$

where $\Phi_{\mathbf{ys}}(j, j)$ and $\Phi_{\mathbf{y}}(j, j)$ are the jth diagonal elements of the covariance matrices $\Phi_{\mathbf{ys}}$ and $\Phi_{\mathbf{y}}$, respectively.

Since the objective is to minimize the scatter within the individual classes, while maximizing the scatter between the classes, the desired number of features with maximum **Q** values are selected as features. The number of features used is very important, since in addition to the computational efficiency, a high dimensionality requires a correspondingly large number of experimental data sets for training the system and developing the classifier. This is because the adequate number of training data sets has to be four or more times the number of features. However, an excessive number of training sets adds to the cost without necessarily improving the system performance. On the other hand, a small number of features may provide an insufficient description of the system.

Once the desired number of features is known, say n, the data sets can then be represented in an n-dimensional feature space to show their relationship to each other. For ease of graphical representation, however, two of the features may be plotted in a two-dimensional feature space (see, for example, Figure 16.6).

Having reduced the data set to a manageable size without significant loss of information, the next step is to identify the source of the data or assign it to a specific class. This step is referred to as pattern classification and is discussed in Section 16.3.

16.3 PATTERN CLASSIFICATION

There are several techniques available for classifying different patterns, but due to space limitations, our discussion focuses on two of the most common methods, pattern recognition and neural network analysis.

16.3.1 Pattern Recognition

The discussion on pattern recognition starts with the classical Bayes decision theory, which is a fundamental statistical approach to pattern classification.

16.3.1.1 Bayes Decision Theory The discussion on Bayes decision theory is approached using a specific example. Let us consider a process which can be in one of s ($s \geq 2$) mutually exclusive states or classes and is monitored at some regular intervals. For $s = 2$, one of the states may be considered to be in control, while the other is out of control. For example, in laser welding, let s_1 indicate the state that the process is in good condition, while s_2 represents a defective weld. Further, let each class or state of the process have an *a priori* probability p_i (*a priori* probability of s_1 is p_1 and that of s_2 is p_2), such that

$$p_1 + p_2 = 1 \tag{16.16}$$

The *a priori* probability of a given state is obtained from previous knowledge of the likelihood that the next observation will belong to that particular state.

Now let the process be monitored by measurement of n-dimensional vectors of observations of an output, say the acoustic emission signal power spectrum, **y**. Since we cannot at any time predict with certainty what the next observation will be, **y** is considered to be a random variable with a distribution that depends on the state. Further, let the class-conditional probability density function for **y** be $p(\mathbf{y}|s_i)$. This is the probability density function for the observation **y** given that the state is s_i (Figure 16.7).

The problem that we are concerned with is to determine, for any observation, **y**, that is made, which one of the s mutually exclusive states has generated **y**, that is, whether the observed **y** is from a good weld or a defective weld. For this, we first take a look at the Bayes rule.

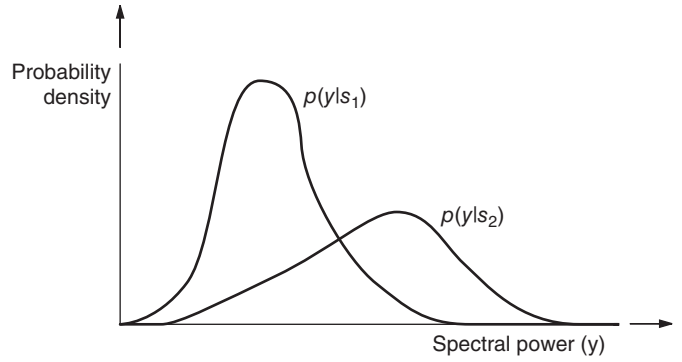


Figure 16.7 Class-conditional probability density functions for two hypothetical classes. Source: From Duda and Hart (1973)/with permission of John Wiley & Sons, Inc.

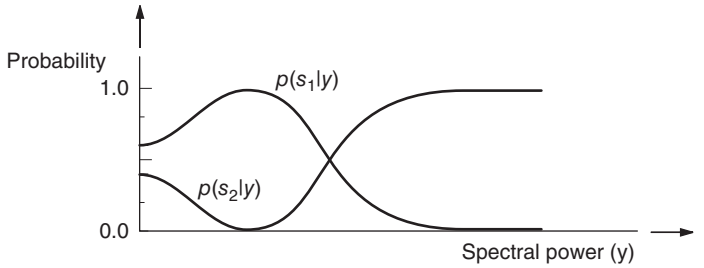


Figure 16.8 *A posteriori* probabilities for two hypothetical classes. Source: From Duda and Hart (1973)/with permission of John Wiley & Sons, Inc.

The Bayes rule expresses the *a posteriori* probability $p(s_i|\mathbf{y})$ (Figure 16.8) in terms of the *a priori* probability p_i and the class-conditional probability density, $p(\mathbf{y}|s_i)$, that is,

$$p(s_i|\mathbf{y}) = \frac{p(\mathbf{y}|s_i)p_i}{p(\mathbf{y})} \tag{16.17}$$

where $p(\mathbf{y}) = \sum_{i=1}^{2} p(\mathbf{y}|s_i)p_i$ for the two-state example being considered.

16.3.1.2 Bayes Decision Rule for Minimum Error With the *a posteriori* probabilities determined using the Bayes rule, they can now be used to classify an observed signal in such a way as to minimize classification error. Now suppose that at some point in the welding process, a single N_m-dimensional observation, $\mathbf{X}' = [x_1, x_2, x_3, \ldots, x_{Nm}]$, is made of say an acoustic emission signal, which after transformation and data reduction becomes $\mathbf{y}' = [y_1, y_2, y_3, \ldots, y_n]$. Let us further suppose we are interested in determining whether the signal is from a good weld (i) or a defective weld (ii), that is, whether the observation belongs to s_1 or s_2. If the observation $\mathbf{y}$ is such that

$$p(s_1|\mathbf{y}) > p(s_2|\mathbf{y}) \tag{16.18}$$

then we assign $\mathbf{y}$ to s_1. This is the Bayes decision rule for minimizing the probability of error. The probability of error associated with a decision is given by:

$$p(\text{error}|\mathbf{y}) = \begin{cases} p(s_1|\mathbf{y}) & \text{if } \mathbf{y} \text{ is assigned to } s_2 \\ p(s_2|\mathbf{y}) & \text{if } \mathbf{y} \text{ is assigned to } s_1 \end{cases} \tag{16.19}$$

Using Eq. (16.17), the decision rule can also be expressed as:

$$p(\mathbf{y}|s_1)p_1 > p(\mathbf{y}|s_2)p_2 \Rightarrow \mathbf{y} \in s_1 \tag{16.20a}$$

or

$$\text{if } \frac{p(\mathbf{y}|s_1)}{p(\mathbf{y}|s_2)} > \frac{p_2}{p_1}, \quad \text{then } \mathbf{y} \in s_1 \tag{16.20b}$$

$p(\mathbf{y}|s_i)$ is the likelihood of s_i with respect to $\mathbf{y}$. And

$$L_R = \frac{p(\mathbf{y}|s_1)}{p(\mathbf{y}|s_2)} \tag{16.20c}$$

is the likelihood ratio. Thus, the Bayes minimum error decision rule decides s_1 if the likelihood ratio exceeds a threshold value, which is independent of $\mathbf{y}$.

16.3.1.3 Discriminant Function Analysis The classification approach presented in the preceding section can be taken one step further by noting that for a multi-class system, the decision rules may be expressed as functions, discriminant functions, $g_i(\mathbf{y})$, with each class or state being associated with one function:

$$g_i(\mathbf{y}), \qquad i = 1, 2, \ldots, s \tag{16.21}$$

The decision rule then becomes

$$\mathbf{y} \in s_i \text{ if } g_i(\mathbf{y}) > g_j(\mathbf{y}) \qquad \text{for all} \quad j \neq i \tag{16.22}$$

This may be represented schematically as in Figure 16.6. For minimum error classification, the discriminant function may be the *a posteriori* probability, $p(s_i|\mathbf{y})$, that is,

$$g_i(\mathbf{y}) = p(s_i|\mathbf{y}) \tag{16.23}$$

Without loss of generality, we replace the function by the logarithm of the *a posteriori* probability, giving

$$g_i(\mathbf{y}) = \log p(s_i|\mathbf{y}) = \log[p(\mathbf{y}|s_i)p_i] \text{ or } g_i(\mathbf{y}) = \log p(\mathbf{y}|s_i) + \log p_i \tag{16.24}$$

The decision rule divides the feature space into decision regions, Ω_i. Contiguous decision regions are bounded by decision boundaries such that

$$g_i(\mathbf{y}) = g_j(\mathbf{y}) \tag{16.25}$$

From Eqs. (16.20) and (16.24), it is evident that the probability density function is necessary for classification. For simplicity, let us consider the case where the system being monitored has a Gaussian distribution. Then

$$p(\mathbf{y}|s_i) = G(\overline{\mathbf{y}}_\mathbf{i}, \mathbf{\Phi}_{\mathbf{y_i}}) = \frac{1}{(2\pi)^{n/2}|\mathbf{\Phi}_{\mathbf{y_i}}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \overline{\mathbf{y}}_\mathbf{i})'\mathbf{\Phi}_{\mathbf{y_i}}{}^{-1}(\mathbf{y} - \overline{\mathbf{y}}_\mathbf{i})\right] \tag{16.26}$$

where $|\mathbf{\Phi}_{\mathbf{y_i}}| = |E[(\mathbf{y} - \overline{\mathbf{y}}_\mathbf{i})(\mathbf{y} - \overline{\mathbf{y}}_\mathbf{i})']|$ is the determinant of the covariance matrix, $\overline{\mathbf{y}}_\mathbf{i} = E[\mathbf{y}_\mathbf{i}]$ and $\mathbf{y}$ is an n-dimensional vector.

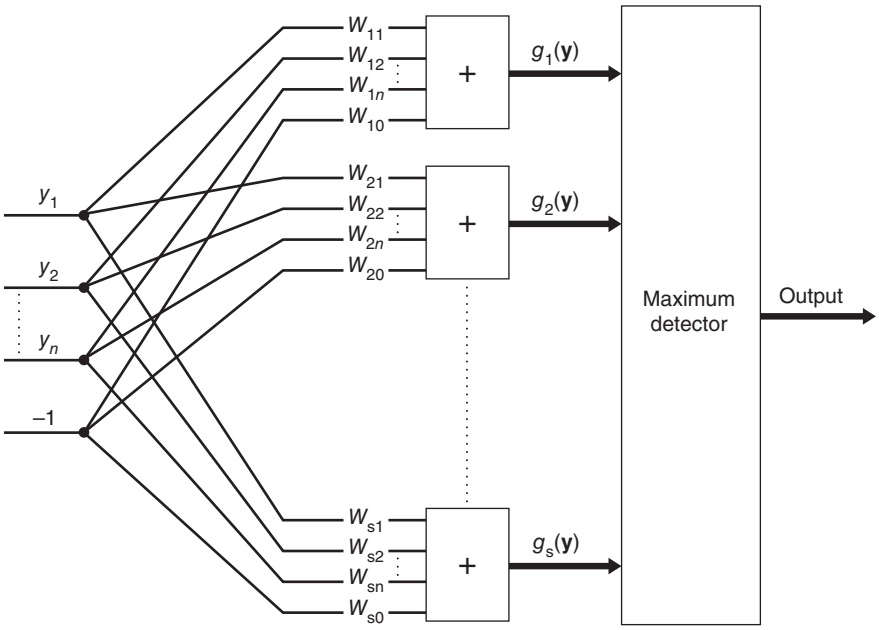


Figure 16.9 An n-dimensional s-class pattern classifier.

Substituting Eq. (16.26) into (16.24) gives

$$g_i(\mathbf{y}) = -\frac{1}{2}(\mathbf{y} - \bar{\mathbf{y}}_\mathbf{i})' \boldsymbol{\Phi}_{\mathbf{y_i}}{}^{-1}(\mathbf{y} - \bar{\mathbf{y}}_\mathbf{i}) - \frac{n}{2}\log 2\pi - \frac{1}{2}\log|\boldsymbol{\Phi}_{\mathbf{y_i}}| + \log p_i \tag{16.27}$$

This is the general form of the discriminant function for a process with normal distribution.

For each sampled signal, the classifier calculates the value of each discriminant function, $g_i(\mathbf{y})$, and assigns the signal to the class with the maximum discriminant function value (Figure 16.9).

16.3.1.4 Least-Squares Minimum Distance Classification Each data set (collectively, the patterns) that is analyzed can be plotted in a feature space using the individual feature components as coordinates. In the preceding classification schemes, it is assumed that the patterns in the feature space are clustered around their respective means, $\bar{\mathbf{y}}_\mathbf{i}$. However, this is not generally the case. Quite often, there is an overlap between the various classes. Thus with a basic linear decision-making process, there can be a substantial amount of misclassification, where signals belonging to one class are erroneously assigned to another. To facilitate clustering of the patterns around their respective means, the patterns may be transformed from the feature space into a decision space. The objective is to map patterns of each class s_i into or to cluster around a predetermined point, $\mathbf{Q}_i$, $i = 1, 2, \ldots, s$ associated with that class.

Thus, for each class, s_i, a point, $\mathbf{Q}_i$, is defined in the decision space. An attempt is then made to transform all patterns belonging to s_i to cluster around $\mathbf{Q}_i$. Since there is still the likelihood of some misclassification even after this step, the transformation is made in such a way as to minimize the decision-making errors that will subsequently result. The transformation matrix, $\mathbf{T}$, is thus selected on basis of the minimum mean square error criterion.

Let us indicate the n-dimensional data set for class s_i in the feature space as $\mathbf{y}_{\mathrm{ij}}, j = 1, 2, \ldots, M_i$. M_i is the number of data sets in class s_i. Further, let this data set be mapped into the predetermined

point $\mathbf{Q}_\mathbf{i}' = [q_1, q_2, \ldots, q_s]$ in the decision space. Also, let the transformed point in the decision space be indicated by $\mathbf{z_{ij}}$:

$$\mathbf{z_{ij}} = \mathbf{T}\mathbf{y_{ij}} \tag{16.28a}$$

Since $\mathbf{z_{ij}}$ will not necessarily be located exactly at $\mathbf{Q_i}$, there will generally be an error ε_{ij}, associated with the transformation of each pattern or data $\mathbf{y_{ij}}$. This is given by:

$$\varepsilon_{\text{ij}} = \mathbf{z_{ij}} - \mathbf{Q_i} = \mathbf{T}\mathbf{y_{ij}} - \mathbf{Q_i} \tag{16.28b}$$

The total mean square error, $\varepsilon_{i,}$ associated with the transformation for class s_i is then

$$\varepsilon_i = \frac{1}{M_i} \sum_{j=1}^{M_i} {\varepsilon_{ij}}^2 \tag{16.26}$$

Substituting from Eq. (16.28b), we get

$$\varepsilon_i = \frac{1}{M_i} \sum_{j=1}^{M_i} [\mathbf{y}_{\mathbf{ij}}' \mathbf{T}' \mathbf{T} \mathbf{y_{ij}} - 2\mathbf{y}_{\mathbf{ij}}' \mathbf{T}' \mathbf{Q_i} + {\mathbf{Q_i}}^2] \tag{16.30}$$

To minimize ε_i with respect to $\mathbf{T}$, we find the gradient:

$$\nabla_{\mathbf{T}} \varepsilon_i = 0$$

which gives

$$\frac{1}{M_i} \sum_{j=1}^{M_i} [\nabla_{\mathbf{T}} (\mathbf{y}_{\mathbf{ij}}' \mathbf{T}' \mathbf{T} \mathbf{y_{ij}}) - \nabla_{\mathbf{T}} (2\mathbf{y}_{\mathbf{ij}}' \mathbf{T}' \mathbf{Q_i}) + \nabla_{\mathbf{T}} ({\mathbf{Q_i}}^2)] = 0 \tag{16.31}$$

Equation (16.31) reduces to

$$\frac{1}{M_i} \sum_{j=1}^{M_i} [2\mathbf{T}(\mathbf{y_{ij}} \mathbf{y}_{\mathbf{ij}}') - (2\mathbf{Q_i} \mathbf{y}_{\mathbf{ij}}') + 0] = 0 \tag{16.32a}$$

or

$$\mathbf{T} \sum_{j=1}^{M_i} (\mathbf{y_{ij}} \mathbf{y}_{\mathbf{ij}}') = \sum_{j=1}^{M_i} (\mathbf{Q_i} \mathbf{y}_{\mathbf{ij}}') \tag{16.32b}$$

which gives the transformation matrix as:

$$\mathbf{T} = \left[\sum_{j=1}^{M_i} (\mathbf{Q_i} \mathbf{y}_{\mathbf{ij}}') \right] \left[\sum_{j=1}^{M_i} (\mathbf{y_{ij}} \mathbf{y}_{\mathbf{ij}}') \right]^{-1} \tag{16.33}$$

In the general case, where each of the classes has a different *a priori* probability p_i, the transformation matrix is given by:

$$\mathbf{T} = \mathbf{G_{Qy}} {\mathbf{G_{yy}}}^{-1} \tag{16.34}$$

where

$$\mathbf{G_{Qy}} = \sum_{i=1}^{s} \sum_{j=1}^{M_i} \frac{p_i}{M_i} (\mathbf{Q_i y'_{ij}}) = E(\mathbf{Qy'}) = \text{cross-correlation matrix}$$

$$\mathbf{G_{yy}} = \sum_{i=1}^{s} \sum_{j=1}^{M_i} \frac{p_i}{M_i} (\mathbf{y_{ij} y'_{ij}}) = E(\mathbf{yy'}) = \text{autocorrelation matrix}$$

The transformation matrix $\mathbf{T}$ obtained in this manner is an $(s \times n)$ matrix.

After transforming the signals into the decision space, appropriate discriminant functions can be developed for classifying the signals. To facilitate representation of the discriminant function in matrix form, the dimensionality of the pattern space is increased by 1 to $(n+1)$ dimensions to account for the threshold in the discriminant function. Thus, $\mathbf{y}$ becomes $\widehat{\mathbf{y}}$, where $\widehat{\mathbf{y}}' = [y_1, y_2, \ldots, y_n, -1]$. In this form, the transformation matrix $\mathbf{T}$ is an $(s \times (n+1))$ matrix.

Applying minimum distance classification to the patterns in the decision space, we determine the distance of the transformed point, $\mathbf{z_{ij}}$, from the preselected point, $\mathbf{Q_i}$, and assign $\mathbf{z_{ij}}$ to the class of the $\mathbf{Q_i}$ to which it is closest. Thus, we have for our discriminant function:

$$d_i^2 = \|\mathbf{z_{ij}} - \mathbf{Q_i}\|^2 = \|\mathbf{z_{ij}}\|^2 - 2\mathbf{Q'_i z_{ij}} + \|\mathbf{Q_i}\|^2 \qquad i = 1, 2, \ldots, s \tag{16.35}$$

The resulting classifier is then known as the least-squares minimum distance classifier. To simplify the analysis, $\mathbf{Q_i}$ can be chosen, without loss of generality, as unit vectors. Thus,

$$\|\mathbf{Q_i}\|^2 = 1$$

and since $\mathbf{z_{ij}}$ is the same for all classes, the classification reduces to determining $\mathbf{Q'_i z_{ij}}$ which is a maximum when d_i^2 is minimum. The discriminant function then becomes

$$g_i(\mathbf{y}) = \mathbf{Q'_i z_{ij}} = \mathbf{Q'_i T y_{ij}} \tag{16.36}$$

and the class with the maximum discriminant function value is selected. As an illustration, for a three-class system, we have

$$\begin{matrix} g_1 \\ g_2 \\ g_3 \end{matrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}_{\mathbf{i}} \mathbf{Ty} = \mathbf{Ty} \tag{16.37}$$

where

$$\mathbf{T} = \begin{bmatrix} w_{11} & w_{12} & \ldots & w_{1n} & w_{10} \\ w_{21} & w_{22} & \ldots & w_{2n} & w_{20} \\ w_{31} & w_{32} & \ldots & w_{3n} & w_{30} \end{bmatrix}_{\mathbf{i}}$$

This gives the discriminant function:

$$g_i(\mathbf{y}) = w_{i1} y_1 + w_{i2} y_2 + \cdots + w_{\text{in}} y_n + w_{i0} \qquad i = 1, 2, \ldots, s \tag{16.38}$$

where w_{ik} is the weighting coefficient of the kth feature for class s_i and w_{i0} is the corresponding threshold.

16.3.1.5 System Training To implement the classifier, it is necessary to determine the elements of the transformation matrix $\mathbf{T}$ or the weighting coefficients w_{ij} of the discriminant function, which

for the linear case, are given by Eq. (16.38). This requires training of the system. For that, data is generated under carefully controlled conditions that produce signals belonging to a particular class without any interference from signals of other classes. For example, if the application is such that cracks can be detrimental to the product, then one may be interested in distinguishing between say, a sound weld and one with cracks.

In training such a system, then it is essential to design experiments that will result in crack formation, without any other discontinuities being formed. The data obtained from such an experiment will then be used to define the discriminant function for the class associated with crack formation. A similar set of experiments is then conducted for a sound weld, and the data obtained used to define the discriminant function for the class related to sound welds. It has to be emphasized that such training will have to be repeated for different conditions.

From this data, the mean, $\overline{\mathbf{y}}_i$, and covariance matrix, $\Phi_{\mathbf{yi}}$, for each class are then obtained and substituted in Eq. (16.27) or (16.34) to give the corresponding discriminant function for each class.

The pattern recognition system is essentially a single-level network. Further improvements in classification may yet be obtained by using a multilayer network or a neural network. The neural network concept is briefly discussed in Section 16.3.2.

16.3.2 Neural Network Analysis

Neural networks are based on a parallel processing architecture where the various processors or nodes are interconnected. Conceptually, they are loosely structured along the basic functioning of the brain and have three components: units (nodes or neurons), links (synapses), and layers (sequencing of groups of neurons). Several neural network models have been developed in the literature. For this discussion, we group them into two broad categories as follows:

1. Standard neural networks.
2. Deep learning networks.

16.3.2.1 Standard Neural Networks Standard neural networks normally have a feedforward architecture where information flow is directional. Here, we discuss a basic model that is suitable for manufacturing applications, with a simplified structure as illustrated in Figure 16.10. The network generally has three or more sections – input layer, hidden layer(s), and output layer. The input layer is the one with no links toward it, and the output layer has no links leading away from it. The network also has the following characteristics:

1. A weight (or synaptic strength) $q(m, n, k)$ for each link (synapse) (m, n) between two nodes: m in layer k and n in layer $k-1$.
2. A residue or threshold $r(m, k)$ for each node m in layer k.
3. A relationship, $I(m, k) = f[q(m, n, k), O(n, k-1), r(m, k)]$, for each node m in layer k:

$$m, n = 0, 1, 2, 3, \ldots \qquad k = 1, 2, 3, \ldots$$

where $I(m, k)$ is the input and $O(m, k)$ is the output for each node m in layer k.

The information available in the network is contained in the magnitude of the links between the nodes. In the general case, each node may be connected to every other node. However, for computational efficiency, a simpler structure involving layers of nodes as illustrated in Figure 16.10 is often used. In this case, each node of a hidden layer (layers other than the input and output layers) is connected to all nodes in the adjacent layers. It receives input from nodes of the preceding layer, and its output becomes part input to each node of the succeeding layer. The performance of

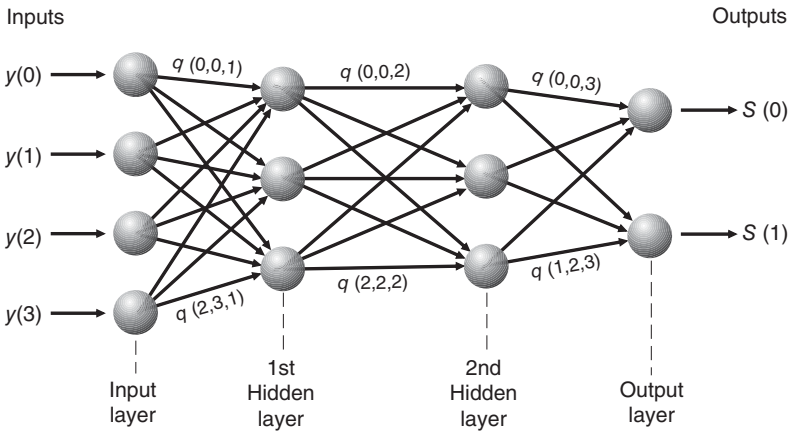


Figure 16.10 Schematic of a standard neural network.

the network depends on the nature of the process being modeled or monitored, the number of nodes used, the number of layers, and the link weights or synaptic strengths, among other things.

If we denote the input to the mth node of the kth layer by $I(m, k)$, and the output of that same node by $O(m, k)$, then the input $I(m, k)$ is given as:

$$I(m,k) = \sum_n [q(m,n,k) \times O(n,k-1)] + r(m,k) \tag{16.39}$$

where $q(m,n,k)$ is the synaptic strength of the link from node n in layer $k-1$ to node m in layer k and $r(m, k)$ is the residue or threshold of node m in layer k.

The inputs to the overall network are denoted as $y(m)$ and the outputs as $s(m)$.

The network is linear when the input to each node is also taken as its output. For a useful network of this form, the input set has to be linearly independent, in which case the network is essentially equivalent to a two-layer network, that is, it can be shown to reduce to a two-layer network. Thus, a nonlinear relationship between the input and output of a node is essential for enhanced performance.

In the general case, the output $O(m, k)$ of each node in layer k is a specific function $f[I(m, k)]$ of the node input $I(m, k)$. A number of functions have been investigated that relate the input and output of a node. One that has been found to be very effective is the monotonically increasing standard sigmoid (or logistic) function:

$$O(m,k) = \frac{1}{1 + \exp[-I(m,k)]} \tag{16.40}$$

This output serves as part input to each node of the $(k+1)$th layer, along with all other outputs of the kth layer. This structure is characterized as a feedforward neural network, and signals can propagate through the network only between adjacent layers, and only in the forward direction. From one layer to another, the signals are modified by the link magnitudes. The basic structure of the network thus consists of three stages, namely the input, hidden layers, and the output, and the input signal propagates through the hidden layers in sequence, being modified at each stage, and finally coming out as the output. The hidden layers process the input information in a way that simulates the system of interest, thereby resulting in an output that is characteristic of the system input.

Network Architecture The simplest network structure would involve direct correspondence between the inputs and outputs, in essence, a two-layer system. However, such simple systems are not as effective as those incorporating hidden layers.

The system performance is strongly influenced by the number of hidden layers and neurons that make up the network. An appropriate number of neurons needs to be used for any given problem. The network will not perform well during the learning stage if the number of neurons is not adequate. In other words, a small number of neurons is unlikely to produce accurate results, and the learning process may be unsuccessful. On the other hand, an unnecessarily large number of hidden layer neurons may increase the number of iterations required to identify the synaptic strengths without necessarily improving the performance. Under such circumstances, several different solutions may be produced, which will make it difficult to come up with the correct generalization for any given input. In other words, a network is less able to generalize for a given input training set as the number of hidden neurons increases. This will cause failure during operation.

One of two methods is usually used in determining the optimal network architecture. In one case, the design starts with a large topology. The system is first trained, and then neurons that do not contribute significantly to the network are eliminated systematically, thereby reducing the network size. A major drawback of this approach is that it is computationally wasteful. The second approach involves starting with a small network and adding one neuron at a time until a solution is obtained. It has a better chance of coming up with the smallest possible network for the task, if the procedure is done right. The drawback is that the network has to be retrained whenever a new neuron is added. The optimum solution is considered to have been achieved when the addition of a neuron does not result in convergence of the solution. This second method is sometimes referred to as the dynamic-node-creation method.

Network Training Determination of the synaptic strengths and activation thresholds (residue) to enable the network perform the mapping of the inputs to the outputs requires training for a supervised network, which is the focus of our discussion. This is accomplished by exposing the system to well-defined states of the process, that is, known outputs corresponding to known inputs. This is repeated for various known input–output sets and the total resulting error determined. The synaptic strengths and activation thresholds are then modified, and the entire process repeated until an error criterion is satisfied. Once the system is subjected to the learning process and the information base established, it can be used for real-time application, even though the training itself is normally off-line. This is essentially similar to the training discussed for the pattern recognition system.

Several methodologies have been developed for training multilayered neural networks. Probably the most common one is the backpropagation method which we discuss in the next seven paragraphs.

Backpropagation is a form of supervised learning, and simply put, it involves searching for the synaptic strength and residue values that minimize the error function given by Eq. (16.41). The gradient of the error with respect to each of the function parameters is calculated and appropriate changes made to the parameters. This is iterated until it converges.

Random values are initially assigned to the link coefficients, $q(m, n, k)$, and residues, $r(m, k)$, to define the network (typically in the range of $+1$ to -1). The weights must not start off with equal values, since it is generally not possible to reach a configuration of unequal weights under such circumstances. One set of inputs is then presented to the network and the outputs generated by the network compared with the corresponding actual or desired outputs. Due to the randomness of the initial network definition, the two sets of outputs will generally be different. The difference or error of each output is then determined. This is repeated for all the input sets available for training. The maximum sum of squares, ε, of the resulting errors is then determined, namely,

$$\varepsilon = Max_{i=1,\ldots,N_{i0}} \sum_{m=1}^{M_0} [R(m,i) - s(m,i)]^2 \tag{16.41}$$

where $R(m, i)$ is the actual or desired mth output of data set i, $s(m, i)$ is the mth network output of data set i, N_{i0} is the number of input–output training sets available, and M_0 is the number of outputs available in a set.

This error measure is compared with a predetermined error threshold, ε_{th}, which serves as a convergence limit and if the measured error is less than the threshold, then the process is terminated. Otherwise, the network link coefficients are recalculated or updated iteratively by backpropagation. In the backpropagation method, the link coefficients preceding each node are determined as follows:

$$q(m,n,k)^{\{j+1\}} = q(m,n,k)^{\{j\}} + \alpha_q \beta(m,k) O(n,k-1) \tag{16.42a}$$

$$r(m,k)^{\{j+1\}} = r(m,k)^{\{j\}} - \alpha_r \beta(m,k) \tag{16.42b}$$

where α_i is the correction gain or learning rate which is positive and usually less than 1. $\beta(m,k)$ is the correction factor or error term for node m. The superscript $j+1$ indicates the next or updated link coefficient, while j represents the current one.

The learning rate α_i can significantly affect the performance of the network. Using a small value for α_i (typically between 0.05 and 0.25) ensures that a solution will be reached. However, that requires a large number of iterations or epochs. The normal approach is to start with a small value of α_i and increase its value as learning proceeds. The higher value will increase the convergence speed. On the other hand, if the value is too high, that may cause the algorithm to oscillate around the minimum.

Convergence can be enhanced by adding a momentum term to smoothen weight changes:

$$\begin{aligned} q(m,n,k)^{\{j+1\}} = {} & q(m,n,k)^{\{j\}} + \alpha_q \beta(m,k) O(n,k-1) \\ & + \gamma_q [q(m,n,k)^{\{j\}} - q(m,n,k)^{\{j-1\}}] \quad 0 < \gamma_q < 1 \end{aligned} \tag{16.43}$$

The goal of the momentum term is to keep the weight changes moving in the same direction. A typical value of the momentum parameter γ_q will be 0.9.

The correction factors are generally given by:

$$\beta(m,k) = O(m,k)[1 - O(m,k)] \sum_k \sum_h q(h,m,k+1)\beta(h,k+1) \tag{16.44a}$$

where k is summed over all layers above the current layer and h is summed over all nodes in those layers. Those for the last or output layer are given by:

$$\beta(m) = s(m)[R(m) - s(m)][1 - s(m)] \tag{16.44b}$$

Once the training is completed, that is, the maximum of the sum of squared errors converges to the threshold error, ε_{th}, then the network is defined for the process in question such that for any process input that is presented to the network, the output predicted by the network should correspond to the actual process output. In a sense, the network is generalized such that it is not only useful for the inputs for which it was trained but also useful for other process inputs. It must be noted, however, that this does not mean the neural network can be used for a completely different process, but rather for processes in the problem domain.

Required Training Sets To ensure adequate training, the network needs to be presented with a minimum number of input training sets, N_{min}, which can be approximately determined using the following relation:

$$N_{\text{min}} = \frac{N_{\text{h}}}{N_0} \left[N_{\text{i}} \log_2 3 + N_0 - \log_2 \left(\frac{N_{\text{h}}}{e} \right) \right] \tag{16.45}$$

where N_{h} is the number of hidden neurons, N_{i} is the number of inputs, and N_0 is the number of outputs. The equation is based on the assumption that the output neurons are binary.

16.3.2.2 Deep Learning Networks Deep learning networks were developed to address some of the problems associated with standard neural networks, such as overfitting and generalization. In a general sense, deep learning architectures consist of several layers, where each layer is a nonlinear processing unit. When applied to neural networks, they enable more complex functions to be represented by increasing the number of layers and units. Just as for standard neural networks, several architectures have been developed for deep learning networks. Here, we outline two of the most common such architectures, the restricted Boltzmann machine (RBM) and deep belief network (DBN).

Restricted Boltzmann Machines The basic structure of the RBM architecture is illustrated in Figure 16.11 and is such that the neurons (units) form a bipartite graph having a layer of visible units and another layer of hidden units. There is full connection between units in the visible and hidden layers. However, there is no connection between units of the same layer.

Training of an RBM starts with a random state in one layer. Gibbs sampling, which is a Monte Carlo Markov Chain (MCMC) technique, is then used to acquire samples from a posterior distribution. Obtaining the states of units in one layer enables all the units in other layers to be updated. The process starts at time 0 with a visible data vector (v_i) which is used to update the hidden data vector (h_j) based on posterior distribution. Now using the hidden vector from time 0, the visible vector is updated at time 1. This is repeated until equilibrium distribution is achieved.

Following that, the weights within the RBM can be determined using the maximum likelihood criterion for the RBM. This involves obtaining the gradient of the training data log-probability to update the weights as follows:

$$\frac{\partial \log p(v)^0}{\partial w_{\text{ij}}} = \langle v_i h_j \rangle^0 - \langle v_i h_j \rangle^\infty \tag{16.46}$$

where p is probability, w_{ij} is the weight or synaptic strength between visible unit i and hidden unit j, $\langle v_i h_j \rangle^0$ and $\langle v_i h_j \rangle^\infty$ indicate the correlations for the visible (v) and hidden (h) units in their initial and equilibrium layer states, respectively. However, training by this method is difficult due to the length of time it takes to reach equilibrium, $\langle v_i h_j \rangle^\infty$ and the high variance associated with it.

A general step of a faster algorithm that is commonly used, the contrastive divergence (CD) algorithm, is outlined in Table 16.1. It replaces the equilibrium term, $\langle v_i h_j \rangle^\infty$, with the first step in the iteration $\langle v_i h_j \rangle^1$ without significant error.

RBMs have found application in a number of areas, including classification, dimensionality reduction, feature learning, and modeling. As such, different versions of the technique have been developed, including robust Boltzmann machines and temperature-based RBMs.

Deep Belief Network In its simplest form, the DBN consists of a number of RBMs which are stacked together, as illustrated in Figure 16.12. It is evident from the figure that each pair of

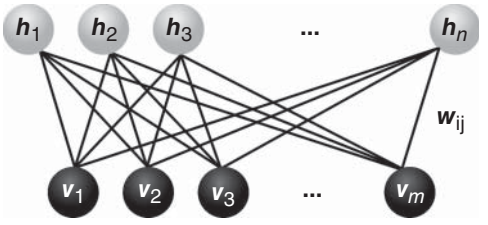


Figure 16.11 Schematic of a restricted Boltzmann machine.

Table 16.1 A general step of the contrastive divergence algorithm for RBMs (Algorithm 1).

Input: RBM($V_1, \ldots, V_m, H_1, \ldots, H_n$), training period T, learning rate ε
Output: The RBM weight matrix w, gradient approximation Δw_{ij}, Δa_i and Δb_j for $i = 1, \ldots, m, j = 1, \ldots, n$

Initialize w with random values distributed uniformly in [0, 1], $\Delta w_{ij} = \Delta a_i = \Delta b_j = 0$

For $\forall\ t = 1$: T do,
  Sample $v_i^{(t)}\ P(v_i\ |h^{(t)})$ when $i = 1, \ldots, m$
  Sample $h_j^{(t+1)}\ P(h_j\ |v^{(t)})$ when $j = 1, \ldots, n$
  For $i = 1, \ldots, m, j = 1, \ldots, n$ do
    $\Delta w_{ij} = \Delta w_{ij} + \varepsilon \times P(H_j = 1|v^{(0)}) \times v_i^{(0)} - \varepsilon \times P(H_j = 1|v^{(T)}) \times v_i^{(T)}$
    $\Delta a_i = \Delta a_i + \varepsilon \times (v_i^{(0)} - v_i^{(T)})$
    $\Delta b_i = \Delta b_i + \varepsilon \times [P(H_j = 1|v^{(0)}) - P(H_j = 1|v^{(T)})]$
    $w = w + \varepsilon \times \Delta w_{ij}$
  End For
End For

Source: Liu et al. (2017)/with permission of Elsevier.

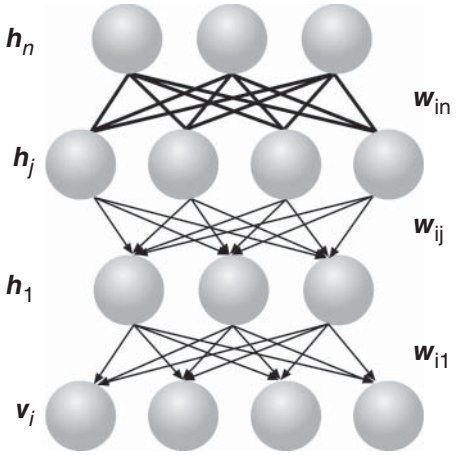


Figure 16.12 Schematic of a deep belief network.

contiguous layers constitutes an RBM, with the first two layers being nondirectional. DBNs are highly suited for problems where the data is unlabeled.

Training starts with the lower RBMs and moves sequentially upward. When features are eventually extracted by the top layers, they are propagated back to the lower layers. The training is a two-stage process, starting with pretraining, followed by fine-tuning. Pretraining involves unsupervised learning from the lower level to the upper level, resulting in feature extraction. The initial weights that are obtained at this stage help improve the network performance, since they are more structured, as opposed to the initial random weights that would otherwise be used. Fine-tuning involves supervised learning that goes from top to bottom to refine the network parameters. An algorithm for the DBN is outlined in Table 16.2.

DBNs have also found application in a number of areas such as image classification, and as a result, different versions of the technique have been developed.

Table 16.2 Deep Belief Network Algorithm (Algorithm 2).

Input: Input visible vector v_{in}, training period T, learning rate ε, number of layers J.
Output: Weight matrix w^i of layer i, $i = 1, 2, \ldots, J-2$.
Initialize w with random values from 0 to 1, $h^0 = v_{in}$; where h^0 denotes the value of units in the input layer. h^i represents the units' value of the i^{th} layer. $layer = 1$;
For $\forall\ t = 1: T$ **do**,
For $layer < J$ **do** Gibbs sampling h^{layer} using $P(h^{layer} \mid h^{layer\text{-}1})$
Computing the CD in Table 16.1, $w(layer)$ is achieved using
$w(t+1) = w(t) + \varepsilon \times \Delta w$
End for
End for

Source: Liu et al. (2017)/with permission of Elsevier.

16.3.3 Sensor Fusion

Whereas most sensors may be suitable for monitoring specific states or detecting specific defects associated with a process, no single sensor can reliably monitor all aspects of the process. Defects that may not be detectable to one sensor under a given set of processing conditions may be detectable to others. Thus by combining multiple sensors, the advantages of each sensor are pooled to form a more reliable system. The drawback, however, is the cost and complexity of the multi-sensor system. Sensor fusion techniques can be categorized into three broad groups:

1. *Low-Level or Data Fusion*: Here, the original data from the different sensors is combined before features are extracted. This type of fusion is suitable for sensors which are identical or whose output data are comparable or can be merged, for example, infrared and ultraviolet output signals. The advantage with this approach is the minimum loss of information.
2. *Medium-Level or Feature Fusion*: With feature fusion, selected features may come from one or more sensors. The data from each sensor is first transformed into the feature space. The features from all the sensors are then concatenated into a single feature vector, before feature selection, and then classification. Since features that are orthogonal in one sensor set (due to the orthogonality of the FFT) may not be orthogonal to features in another sensor set, the resulting features may need to be processed by an orthogonal transform before pattern classification. The advantage with this approach is that it permits fusion of data from sensors which are not commensurate.
3. *High-Level or Decision Fusion*: This invokes a voting system, with the data from each sensor being processed independently until a decision is made. The decisions from each sensor are then weighted and then combined to give the overall results through voting.

Whenever possible, it is generally recommended that the lowest level of fusion possible should be used, since it results in minimal loss of information. Hybrids or combinations of the different fusion methods may also be used.

16.3.4 Time–Frequency Analysis

The Fourier transform, as defined in Section 16.1.1, requires knowledge of the signal over the infinite range, that is, from $-\infty$ to ∞. This mathematical expression only represents the real physical situation for specific cases such as steady state or stationary signals. For transients, this represents

a distortion of the physical situation, even though it is mathematically correct since the results of the analysis are such that the signal vanishes outside its domain. Furthermore, it does not properly represent localized information, for example, spikes and high-frequency bursts. Such anomalies are addressed using time–frequency analysis, which is suitable for nonstationary signals since its frequency characteristics evolve with time. It enables the frequency content of a signal to be obtained locally in time (or at any "instant" in time).

Several techniques have been developed for time–frequency analysis. Perhaps the simplest one is a direct extension of the regular Fourier transform, the STFT, also known as windowed Fourier transform.

16.3.4.1 Short-Time Fourier Transform This technique involves taking the Fourier transform of short segments or windows of the signal one after another and concatenating the resulting spectrograms. Application of the window is illustrated in Figure 16.13. In essence, well-localized slices of the function are taken one at a time by sliding the window function along the time axis and taking snapshots of the spectrograms. The result is a three-dimensional plot of the amplitude, frequency, and time. Quite often a two-dimensional plot in the time–frequency domain (a phase space) is used, with amplitude being represented by a gray scale or color coding.

The STFT of a signal $f(t)$ can be expressed as:

$$\text{STFT}(\omega, \tau) = \int_{-\infty}^{\infty} f(t)\omega(t-\tau)e^{-i\omega t}dt \tag{16.47}$$

where $\omega(t)$ is the window function and $\omega(t-\tau)$ gives the region around the time of interest τ over which the spectrum is obtained.

The discrete form of the transform is given by:

$$\text{STFT}\,(\omega, k) = \sum_{j=-\infty}^{\infty} x(j)\omega(j-k)e^{-i\omega j} \quad k \text{ is discrete} \tag{16.48}$$

The signal $f(t)$ is suppressed by the window, $\omega(t-\tau)$ outside its domain, such that the Fourier transform results in a local spectrum. Several forms of the window function have been developed. The simplest case is that of the rectangular window function. However, as we know from

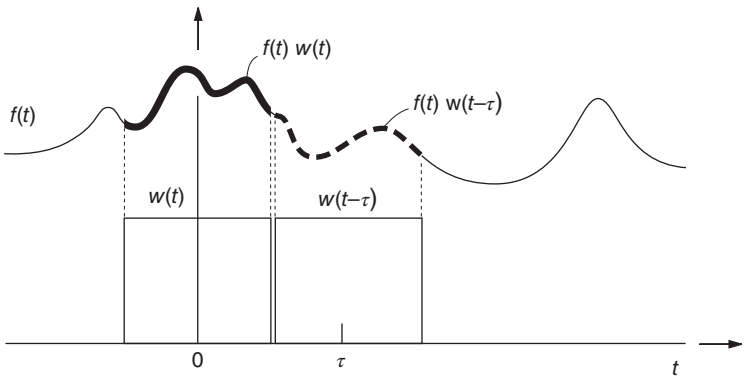


Figure 16.13 Windowed function to be used for the short-time Fourier transform.

Section 16.1.6, the rectangular window introduces side lobes into the frequency domain, and that can result in leakage. When a Gaussian window is used, then we have the Gabor transform:

$$G_f(\omega, \tau) = \int_{-\infty}^{\infty} f(t) e^{-\pi(t-\tau)^2} e^{-i\omega t} dt \tag{16.49}$$

From Section 16.1.2, we know that the resolution of the STFT is determined by the width of the window function. The wider the window, the better the frequency resolution. However, the time resolution then deteriorates. On the other hand, a narrower window results in better time resolution and poorer frequency resolution. That is one drawback of the STFT technique. To address this issue, other techniques have been developed for time–frequency analysis. The major ones include wavelet transforms and time–frequency distributions. In the next section, we take a brief look at the wavelet transform.

16.3.4.2 Wavelet Transforms The time and frequency resolution problem associated with the STFT is overcome using wavelets, since it is possible to adapt the window length of wavelets to the frequency. As a result, wavelets are able to track rapid variations in frequency. They are thus capable of giving good time resolution for appropriate regions of the signal (high-frequency segments) and likewise, good frequency resolution for the low-frequency segments. A wavelet, $\psi(t)$, is essentially a family of functions known as the mother wavelet which is characterized by position and frequency, and whose average value is zero:

$$\int_{-\infty}^{\infty} \psi(t) dt = 0 \tag{16.50}$$

Associated with the mother wavelet are scaled and translated versions, often known as daughter wavelets. In essence, wavelets segment data into different frequency components and analyze each component with a resolution that is commensurate with its scale. The wavelet $\psi(t)$ is expressed as:

$$\psi_{(a,b)}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right) \tag{16.51}$$

where a is a dilation or scale parameter, known as the scalogram, that governs the frequency or width of the wavelet and b indicates the wavelet position. Thus, a variation in b implies a translation in time.

The factor $1/\sqrt{|a|}$ ensures that all scaled functions have the same energy. The smaller the absolute value of a, the higher the frequency content of the wavelet ψ.

A wavelet transform essentially represents a function by wavelets, and the transform of a function $f(t)$ with respect to wavelet $\psi(t)$ is given by:

$$\mathrm{WF}(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(t) \widetilde{\psi}\left(\frac{t-b}{a}\right) dt \tag{16.52}$$

where $\widetilde{\psi}(t)$ is the complex conjugate of $\psi(t)$.

The discrete wavelet transform (DWT) can be viewed simply as the wavelet transform of a sampled sequence, $x_j = x(j\delta\tau)$, which for a unit sampling period, can be expressed as:

$$\mathrm{DWT}\,[x_j; a, b] = \sum_j x(j) \widetilde{\psi}_{a,b}(j) \tag{16.53}$$

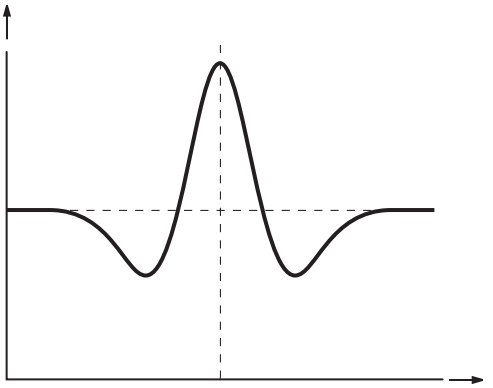


Figure 16.14 The Mexican hat wavelet.

Several types of wavelet functions have been developed. Among the discrete wavelets are the Haar and Daubechies wavelets, and among the continuous ones are the Morlet and Mexican hat wavelets. The Haar wavelet is given by:

$$\psi(t) = \begin{cases} 1 & 0 \le t < \frac{1}{2} \\ -1 & \frac{1}{2} \le t < 1 \\ 0 & \text{otherwise} \end{cases} \tag{16.54a}$$

with the scaling function:

$$\varphi(t) = \begin{cases} 1 & 0 \le t < 1 \\ 0 & \text{otherwise} \end{cases} \tag{16.54b}$$

Its main disadvantage is that it is not particularly good for approximating smooth functions, since it is not continuous and therefore not differentiable. The Mexican hat wavelet is given by:

$$\psi(t) = \frac{1}{\sqrt{2\pi}\sigma^3}\left[1 - \left(\frac{t}{\sigma}\right)^2\right] e^{-\frac{1}{2}\left(\frac{t}{\sigma}\right)^2} \tag{16.55}$$

It is the second derivative of a Gaussian function that is normalized (Figure 16.14).

16.3.5 Applications in Manufacturing

The time–frequency technique that is used for a given manufacturing process depends on the process characteristics. For processes such as lathe turning or welding (including laser) of a long prismatic part, the STFT used with an appropriate window function should be adequate for process monitoring since the processes are essentially stationary. On the other hand, for processes such as milling, some sheet metal forming or bulk deformation operations, spot welding, or processes involving pulsed laser beams, it will be more appropriate to use wavelet analysis or time–frequency distributions due to the nonstationarity of the processes.

16.4 SUMMARY

Sensor outputs usually have to be analyzed to extract relevant information about the processes that they monitor. One approach is to classify the process output into different states, and this involves three steps: sampling (digitizing), feature extraction, and classification.

Sampling requires certain precautions to avoid contaminating the signal. According to the sampling theorem, the theoretical sampling rate has to be at least twice (in practice about 5–10 times) the highest frequency in the signal. Aliasing results if the sampling rate is less. Frequencies above the folding frequency are then folded back into the frequency range below the folding frequency. Using a finite length of sampled data is equivalent to multiplying the signal by a rectangular window, which introduces high-frequency side lobes into the signal, creating potential aliasing problems.

Feature extraction involves transformation of the signal, along with a reduction in the data size. Among the common transforms are the Fourier, Walsh–Hadamard, Haar, and Karhunen–Loeve transforms, with the Fourier transform being the most commonly used one. The discrete form of the Fourier transform has several advantages over the analog transform: high signal-to-noise ratio; high resolution; no drift; and greater flexibility in analysis. However, again certain precautions need to be taken to get meaningful results.

Data reduction is used to reduce the initial data size to increase computation speed, with minimal loss of information. Among the most common criteria used for data reduction are the variance criterion (which eliminates the feature components with the smallest variances) and the Fisher criterion (which selects the features that minimize scatter of data within individual classes, while increasing the scatter between classes).

Two of the most common methods for distinguishing between different signal sources are pattern recognition and neural network analysis. Bayes' decision theory provides a fundamental basis to pattern recognition. It expresses the *a posteriori* probability in terms of the *a priori* probability and the class-conditional probability density. The decision can be based on minimum error or minimum risk. The resulting decision rules can be expressed as discriminant functions.

Neural networks represent the classifier by nodes which are interconnected, with information available in the network being contained in the magnitude of the links between the processors. Deep learning networks were introduced to overcome some of the problems associated with traditional networks, and two of the most common architectures, RBMs and DBNs are outlined. Each classifier needs to be trained to determine the coefficients and thresholds of discriminant functions, or the link magnitudes of neural networks before being used to monitor the process.

Since the traditional Fourier transform is unable to extract information on the time variation of a signal, it is mainly suitable for stationary processes. For nonstationary processes, it is necessary to use time–frequency methods. The simplest one is the STFT, which is essentially an extension of the traditional Fourier transform. Here, a series of transforms are obtained by sliding a window along the time domain. However, the frequency resolution tends to be poor when the time domain window is short. The use of wavelets and time–frequency distributions enable high-frequency resolutions to be obtained over short time periods, and vice versa.

APPENDIX 16.A

List of symbols used in the chapter

Symbol	Parameter	Units
$a(\omega)$	Amplitude spectrum	—

Symbol	Parameter	Units
$A(\omega)$	Real component of $g(\omega)$	—
$B(\omega)$	Complex component of $g(\omega)$	—
c_i	Constants	—
$f(t)$	Signal function	—
$g_i(\mathbf{y})$	Discriminant function	—
$g(\omega)$	Fourier transform of $f(t)$	—
$\mathbf{G_{Qy}}$	Cross-correlation matrix	—
$\mathbf{G_{yy}}$	Autocorrelation matrix	—
h_{j}	Hidden data vector	
i, j, k, l, m	Integers	—
$I(m, k)$	Input to the mth node of the kth layer	—
J	Number of layers	
K	Number of different configurations that a network can have	
L_{R}	Likelihood ratio	—
M_j	Number of patterns (data sets) in class s_j class	—
M_0	Number of outputs available in a set	—
N_{h}	Number of hidden neurons	
N_i	Number of inputs	
N_m	Number of sampled data points in the time series	—
N_{o}	Number of outputs	
N_{io}	Number of input–output training sets available	—
$O(m, k)$	Output of the mth node of the kth layer	—
p	Probability	—
p_i	*a priori* probability of class s_i	—
$p(\mathbf{y}\|s_i)$	Class-conditional probability density function for $\mathbf{y}$	—
$p(s_i\|\mathbf{y})$	*a posteriori* probability	—
$q(\omega)$	Power spectrum	—
$q(m, n, k)$	Magnitude of link from node n in layer $k-1$ to node m in layer k	—
$\mathbf{Q}$	Feature selection criterion	—
$\mathbf{Q}_i$	Cluster point i in decision space	—
$r(m, k)$	Residue or threshold of node m in layer k	—
$R(m, i)$	The actual or desired mth output of data set I	—
s	Number of classes	—
$s(m, i)$	The mth network output of data set i	—
s_i	ith class	—
T	Training period	
$\mathbf{T}$	Orthogonal transform	—
v_i	Visible data vector	
w_{ij}	Weight or synaptic strength between visible unit i and hidden unit j	—
$x(j)$	Signal time series	—
$\mathbf{x}$	Signal time series vector	—
$\hat{\mathbf{x}}$	Estimate of $\mathbf{x}$	—
$y(k)$	Discrete Fourier transform of $x(j)$	—
$\mathbf{y}$	Fourier transform of $\mathbf{x}$	—
$\mathbf{y_{ij}}$	jth pattern in class s_i	—
$\overline{\mathbf{y}}_{\mathbf{ij}}$	Feature mean for class i	—

Symbol	Parameter	Units
$\bar{\mathbf{y}}$	Overall system mean	—
$\bar{\mathbf{y}}_\mathbf{i}$	Mean of $\mathbf{y_i}$	—
$z(j)$	Shifted time series	—
$\mathbf{z_{ij}}$	Transformation of $\mathbf{y_{ij}}$ in decision space	—
α_i	Correction gain or learning rate	—
$\beta(m, k)$	Correction factor or error term for node m	—
$\delta\tau$	Sampling period	s
$\Delta\tau$	Signal duration	s
ε	Learning rate	
ε	Mean square error	—
λ_i	Lagrange multiplier or eigenvalue	—
ν_a	Aliasing frequency	Hz
ν_f	Folding frequency	Hz
ν_h	Highest frequency in a signal	Hz
ν_s	Sampling frequency (sampling rate)	Hz
ω	Angular frequency	rad/s
Ω_i	Region in the domain of $\mathbf{y}$ for which $\mathbf{y} \in s_i$	—
$\mathbf{\Phi_x}$	Covariance matrix of $\mathbf{x}$	—
$\mathbf{\Phi_y}$	Covariance matrix of $\mathbf{y}$	—
$\mathbf{\Phi_{yi}}$	Scatter within each class	—
$\mathbf{\Phi_{ys}}$	Scatter between individual classes	—
σ^2	Variance	—

PROBLEMS

16.1. Given a system with a normal conditional probability density function,

$$p(\mathbf{y}|s_i) = G(\bar{\mathbf{y}}_\mathbf{i}, \mathbf{\Phi}_{\mathbf{y_i}}) = \frac{1}{(2\pi)^{n/2}|\mathbf{\Phi}_{\mathbf{y_i}}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \bar{\mathbf{y}}_\mathbf{i})'\mathbf{\Phi}_{\mathbf{y_i}}{}^{-1}(\mathbf{y} - \bar{\mathbf{y}}_\mathbf{i})\right]$$

with a different covariance matrix $\mathbf{\Phi_{yi}}$ for each class, obtain a general expression for the discriminant function, $\mathrm{g_i(y)}$, for classifying the system using the Bayes decision rule and the logarithm of the a posteriori probability, if the a priori probability is $\mathrm{p_i}$.

16.2. Derive the relationship

$$\mathbf{x} = [\mathbf{A}'\mathbf{A}]^{-1}\mathbf{A}'\mathbf{y}$$

given that $\mathbf{y} = \mathbf{Ax}$ and $\nabla_\mathbf{x}\|\mathbf{Ax} - \mathbf{y}\|^2 = 0$.

Also, **A, x, and y** are $(n \times m)$, $(m \times 1)$, and $(n \times 1)$, respectively, with $[\mathbf{A}'\mathbf{A}]$ being nonsingular.

16.3. For the following two-dimensional training set

$$S_1 = \begin{bmatrix}1\\3\end{bmatrix}, \quad \begin{bmatrix}-1\\4\end{bmatrix}, \quad \begin{bmatrix}-3\\4\end{bmatrix}, \quad \begin{bmatrix}0\\2\end{bmatrix}, \quad \begin{bmatrix}-4\\2\end{bmatrix}$$

$$S_2 = \begin{bmatrix}5\\6\end{bmatrix}, \quad \begin{bmatrix}4\\5\end{bmatrix}, \quad \begin{bmatrix}6\\6\end{bmatrix}, \quad \begin{bmatrix}8\\8\end{bmatrix}, \quad \begin{bmatrix}7\\6\end{bmatrix}$$

$$S_3 = \begin{bmatrix}5\\0\end{bmatrix}, \quad \begin{bmatrix}6\\1\end{bmatrix}, \quad \begin{bmatrix}7\\-1\end{bmatrix}, \quad \begin{bmatrix}8\\0\end{bmatrix}, \quad \begin{bmatrix}9\\2\end{bmatrix}$$

if the three classes have equal *a priori* probabilities, determine

(a) The discriminant functions for the classifier.
(b) The equations of the decision boundaries.

16.4. Consider $(n \times 1)$ vectors $\mathbf{x}$ and $\mathbf{y}$, and an $(n \times n)$ matrix $\mathbf{A}$.

(a) Find F if

$$F = \nabla_{\mathbf{A}}[\mathbf{y}'\mathbf{y}]$$

(b) Determine

$$\nabla_{\mathbf{A}}[\mathbf{x}'\mathbf{A}'\mathbf{y}]$$

16.5. Show that the following transformation matrix is orthonormal:

$$\mathbf{T} = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & 3 \\ -3 & 1 \end{bmatrix}$$

16.6. For the following two-dimensional training set

$$S_1 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \quad \begin{bmatrix} -1 \\ 5 \end{bmatrix}$$

$$S_2 = \begin{bmatrix} 4 \\ 7 \end{bmatrix}, \quad \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$

$$S_2 = \begin{bmatrix} 3 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 5 \\ -1 \end{bmatrix}$$

if the three classes have equal *a priori* probabilities, determine

(a) The discriminant functions for the classifier.
(b) The equations of the decision boundaries.

16.7. Consider the one-dimensional form of the Cauchy distribution, and let the class-conditional density be expressed as:

$$p(y|s_i) = \frac{1}{\pi a}\left[\frac{1}{1 + \left(\dfrac{y - b_i}{a}\right)^2}\right], \qquad i = 1, 2$$

Assuming equal *a priori* probabilities, show that the minimum probability of error is given by:

$$p(\text{error}) = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\left|\frac{b_2 - b_1}{2a}\right|$$

16.8. Given $\mathbf{x}' = [x_1, x_2, \ldots, x_n]$, and

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

show that

$$\nabla_{\mathbf{A}}[\mathbf{x}'\mathbf{A}'\mathbf{A}\mathbf{x}] = 2\mathbf{A}(\mathbf{x}\mathbf{x}')$$

16.9. Obtain the covariance matrix $\Phi_{\mathbf{x}}$ of the following data set:

$$\mathbf{x}_1 = \begin{bmatrix}0\\3\end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix}1\\3\end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix}1\\2\end{bmatrix}, \quad \mathbf{x}_4 = \begin{bmatrix}1\\-1\end{bmatrix}, \quad \mathbf{x}_5 = \begin{bmatrix}2\\3\end{bmatrix}$$

16.10. Consider a two-dimensional training set to be as follows:

$$S_1 = \begin{bmatrix}-2\\3\end{bmatrix}, \quad \begin{bmatrix}-3\\3\end{bmatrix}, \quad \begin{bmatrix}-4\\3\end{bmatrix}$$

$$S_2 = \begin{bmatrix}6\\6\end{bmatrix}, \quad \begin{bmatrix}6\\7\end{bmatrix}, \quad \begin{bmatrix}7\\5\end{bmatrix}$$

(a) Obtain the discriminant functions for the classifier, assuming equal *a priori* probabilities, and using the least-squares minimum distance classifier.
(b) Determine the equation of the decision boundary.

16.11. A multivariate normal density has $\sigma_{ij} = 0$ and $\sigma_{ii} = \sigma_i{}^2$. Show that the normal density function for this case is given by:

$$p(\mathbf{x}) = \frac{1}{\prod_{i=1}^{n}\sqrt{2\pi}\sigma_i} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right]$$

where n is the rank or dimensionality of x, Σ indicates summation, and Π indicates product.

16.12. Show that for an orthonormal transformation matrix $\mathbf{T}$ and a real symmetric matrix $\mathbf{A}$,

$$\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{T}\mathbf{A}\mathbf{T}') = \sum a_{jj}$$

where a_{jj} are the diagonal elements of $\mathbf{A}$.

BIBLIOGRAPHY

Ahmed, R. and Rao, K.R. (1975). *Orthogonal Transforms and Digital Signal Processing*. Berlin: Springer-Verlag.

Akay, M. (1998). *Time Frequency and Wavelets in Biomedical Signal Processing*. Piscataway, NJ: IEEE Press.

Anderson, K., Cook, G.E., Karsai, G., and Ramaswamy, K. (1988). Artificial neural networks applied to arc welding process modeling and control. *IEEE Transactions on Industry Applications* 26: 824–830.

Bergland, G.D. (1969). A guided tour of the fast Fourier transform. *IEEE Spectrum* 6: 41–51.

Boashash, B. (1992). *Time-Frequency Signal Analysis - Methods and Applications*. Melbourne, Australia: Longman Cheshire.

Brooks, R.R. (1997). *Multi-Sensor Fusion: Fundamentals and Applications with Software*. Upper Saddle River, NJ: Prentice Hall.

Carpenter, G.A., Grossberg, S., and Reynolds, J.H. (1991). ARTMAP: supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks* 4: 565–588.

Carpenter, G.A., Grossberg, S., Markuzon, N. et al. (1992). Fuzzy ARTMAP: a neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks* 3 (5): 698–713.

Choi, H.I. and Williams, W.J. (1989). Improved time-frequency representation of multi component signals using exponential kernels. *IEEE Transactions on Acoustics Speech Signal Processing* 37 (6): 862–871.

Cohen, L. (1966). Generalized phase-space distribution functions. *Journal of Mathematical Physics* 7: 781–786.

Daubechies, I. (1990). The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory* 36: 961–1005.

Duda, R.O. and Hart, P.E. (1973). *Pattern Classification and Scene Analysis*. New York: Wiley.

Fausett, L. (1994). *Fundamentals of Neural Networks*. Englewood Cliffs, NJ: Prentice Hall.

Fischer, A. and Igel, C. (2014). Training restricted Boltzmann machines: an introduction. *Pattern Recognition* 47: 25–39.

Freeman, J.A. and Skapura, D.M. (1991). *Neural Networks – Algorithms, Applications, and Programming Techniques*. Reading, MA: Addison-Wesley.

Fukunaga, K. (1972). *Introduction to Statistical Pattern Recognition*. Orlando: Academic Press, Inc.

Ganssle, G. (2018). Neural networks. *The Leading Edge* 616–619.

Graupe, D. (2007). *Principles of Artificial Neural Networks*, 2e. Singapore: World Scientific.

Hall, D.L. (1992). *Mathematical Techniques in Multisensor Data Fusion*. Boston, MA: Artech House.

Haykin, S. (1994). *Neural Networks: A Comprehensive Foundation*. New York: Macmillan.

Hess-Nielson, N. and Wickerhauser, M.V. (1996). Wavelets and time-frequency analysis. *Proceedings of the IEEE* 84 (4): 523–540.

Hinton, G.E. (2002). Training products of experts by minimizing contrastive divergence. *Neural Computation* 14: 1771–1800.

Hinton, G.E., Osindero, S., and Teh, Y.W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation* 18 (7): 1527–1554.

Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences* 79: 2554–2558.

Hush, D.R. and Horne, B.G. (1993). Progress in supervised neural networks. *IEEE Signal Processing Magazine* 10 (1): 8–39.

Kannatey-Asibu, E. Jr., and Emel, E. (1987). Linear discriminant function analysis of acoustic emission signals for tool condition monitoring. *Journal of Mechanical Systems and Signal Processing* 1 (4): 333–347.

Klein, L.A. (2004). *Sensor and Data Fusion*. Bellingham, WA: SPIE Press.

Lippmann, R.P. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine* 4 (2): 4–22.

Liu, W., Wang, Z., Liu, X. et al. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing* 234: 11–26.

Mallat, S. (1999). *A Wavelet Tour of Signal Processing*, 2e. San Diego, CA: Academic Press.

Mertins, A. (1999). *Signal Analysis*. New York: Wiley.

Muller, B., Reinhardt, J., and Strickland, M.T. (1995). *Neural Networks: An Introduction*. New York, NY: Springer-Ver1ag Berlin Heidelberg.

Pao, Y.-H. (1989). *Adaptive Pattern Recognition and Neural Networks*. Reading, MA: Addison-Wesley.

Poggio, T. and Girosi, F. (1990). Networks for approximation and learning. *Proceedings of IEEE* 78: 1481–1497.

Rumelhart, D.E., Hinton, G.E., and Williams, R.J. (1986). Learning representations by back-propagating errors. *Nature* 323: 533–536.

Rumelhart, D. and McClelland, J. (1986). *Parallel Distributed Processing*, vol. 1. Cambridge, MA: MIT Press.

Stanley, W.D. (1975). *Digital Signal Processing*. Reston, VA: Reston Publishing.

Sun, A., Kannatey-Asibu, E. Jr., and Gartner, M. (2002). Monitoring of laser weld penetration using sensor fusion. *Journal of Laser Applications* 14 (2): 114–121.

Sun, A., Kannatey-Asibu, E., Gartner, M., and Williams, W. (2001). Time-frequency analysis of laser weld signature. *Proceedings of SPIE Advanced Signal Processing Algorithms, Architectures, and Implementations XI* 4474: 103–114.

Yildirim, I. (2012). Bayesian Inference: Gibbs Sampling, Technical Note. University of Rochester.

17 Laser Safety

Lasers, like most other equipment used in processing materials, require caution for safe and effective utilization. In this chapter, the characteristics of the basic laser hazards that need to be understood for safe use of lasers are outlined, focusing on their application in materials processing. The discussion starts with the hazards associated with lasers. This is followed by the principal classification of lasers. Finally, the basic steps that need to be taken to prevent laser accidents are presented.

17.1 LASER HAZARDS

There are two principal forms of hazard associated with lasers:

1. Radiation-related hazards.
2. Nonbeam hazards.

17.1.1 Radiation-Related Hazards

Assessment of the radiation-related hazards associated with a laser or laser system is normally based on three primary issues:

1. The capacity of the laser and its ability to cause injury.
2. The environment in which the laser is to be used.
3. The personnel who will be associated with the laser (using or being exposed to it).

The damaging effects of a laser beam result primarily from:

1. Its highly collimated nature which enables it to be focused to very small sizes.
2. Its monochromaticity that can result in selective interactions with human tissue.

Since different wavelengths interact differently with tissue, the type of injury that results normally depends on the type of laser being used. Injuries that result from radiation-related hazards are primarily to the human skin and eye, and given the same conditions, the seriousness of the injury depends on where it occurs. Whereas a burn on the hand may leave a permanent scar, an equivalent damage to the eye may cause blindness. The primary factors that determine the extent of damage sustained are as follows:

1. Beam power density.

Principles of Laser Materials Processing: Developments and Applications, Second Edition. Elijah Kannatey-Asibu, Jr.
© 2023 John Wiley & Sons, Inc. Published 2023 by John Wiley & Sons, Inc.
Companion Website: www.wiley.com/go/Kannatey-Asibu/PrinicplesLaserMaterialsProcessing

2. Beam wavelength.
3. Irradiated area and its absorption characteristics.
4. Duration of exposure.

17.1.1.1 Mechanisms of Laser Damage To understand how damage may be caused, the principal mechanisms responsible for laser damage to human tissue are briefly outlined. These are as follows:

1. Photothermal effects.
2. Photochemical effects.
3. Photoacoustic effects.

Photothermal Effects With this mechanism, the radiation absorbed by the tissue increases the tissue's temperature until damage or a laser burn occurs. The process depends more on the rate at which incident energy is absorbed, that is, the power, rather than the total amount of energy absorbed. Thus, injuries that immediately follow exposure are more likely to be due to photothermal effects. Even though most cells can survive temperatures as high as 45 °C, cells begin to die at temperatures of 60 °C, and tissue evaporation would normally occur at about 350 °C. Photothermal effects can thus result in immediate and permanent damage to body tissue (skin or eye).

Photochemical Effects Photochemical effects depend more on the total amount of energy absorbed by the tissue rather than on the rate at which it is absorbed. Its effect is thus cumulative. Thus, even though the power levels may not be sufficient to cause thermal damage, over long periods of exposure, the effect can be damaging. The mechanism involves changes in the chemical structure of the underlying tissue. For example, the creation of a skin tan by absorption of ultraviolet radiation from the sun is photochemical in nature. Even though such photochemical effects may be reversible once absorption of the radiation is stopped, cumulative effects may result in longer-term damage.

As a result of this being an energy-dependent process, the photon energy plays a significant role. And since the photon energy increases with decreasing wavelength, ultraviolet beams are more damaging than visible radiation, even though visible radiation can affect sensitive tissue such as the retina of the eye as a result of prolonged exposure. Infrared radiation generally does not cause damage by the photochemical mechanism because of the low photon energy.

Photoacoustic Effects Photoacoustic effects are generally associated with ultrashort duration pulses (typically less than 10 μs) of high energy which cause rapid expansion of the tissue on which they are incident. This generates an acoustic shock wave in the tissue that may cause mechanical disruption of cellular structures. This process is energy-dependent.

The major forms of radiation-related hazards, eye and skin hazards, are now discussed.

17.1.1.2 Major Hazards

Eye Hazards For effective protection of the eye against laser hazards, it is necessary to have a basic understanding of the eye structure and how it interacts with incident radiation. Figure 17.1 shows the basic components of the eye. Light (radiation) enters the eye through the cornea, passes through the pupil which is the opening within the iris, and is focused on the retina by the lens, which has an effective focal length of 17 mm. The diameter of the pupil can be as small as 2 mm and as large as 7 mm. Under normal circumstances, since light enters the eye from several different directions, an image is formed over the entire retina (Figure 17.2a). Thus, the power density of the light incident on the retina is relatively low. However, when a collimated beam is incident on the lens, it is focused onto a small spot (Figure 17.2b) of diameter about 20 μm on the retina. This results in a magnification

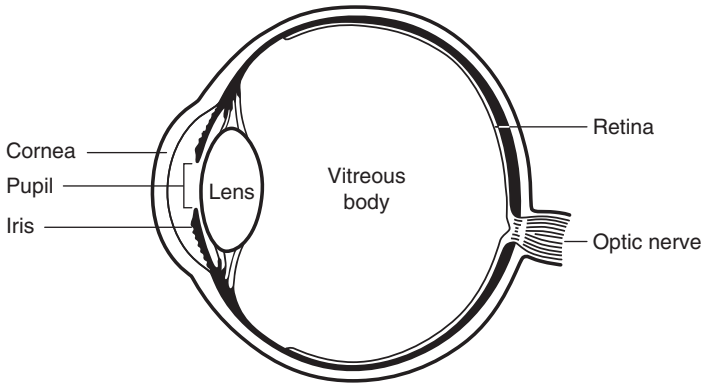


Figure 17.1 Schematic of the basic components of the human eye.

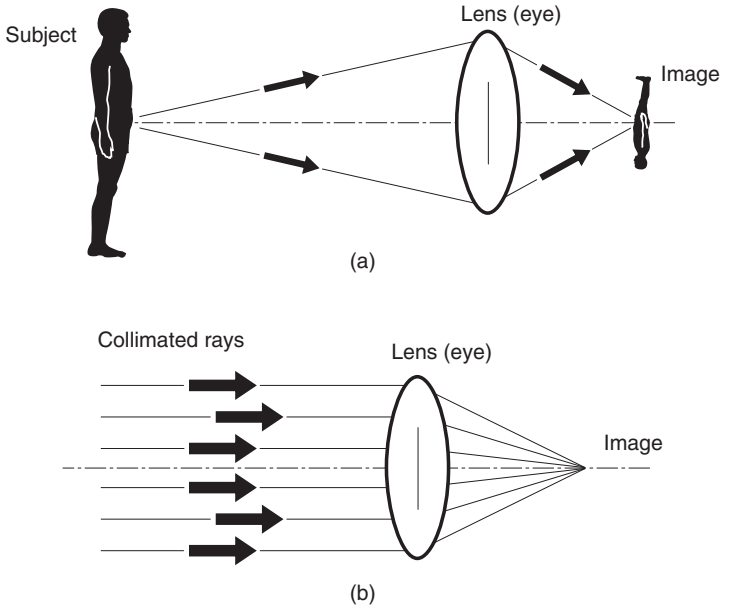


Figure 17.2 Focusing of (a) an object and (b) a collimated beam.

of the light intensity of about 10^5 on the retina. Thus, radiation which might appear to be of relatively low intensity on the outside could cause significant damage on the retina.

Figure 17.3 summarizes the absorption characteristics of the different components of the eye. All radiation in the wavelength range below 0.3 μm (ultraviolet) and above 3 μm (far infrared) is absorbed by the cornea. Radiation that falls between 0.3 and 0.4 μm (near-ultraviolet) and also between 1.4 and 3 μm (mid-infrared) is absorbed by the lens. The remaining wavelength range between 0.4 and 1.4 μm is focused on the retina. This wavelength range is known as the ocular focus or retinal hazard region. It must be noted that part of the radiation in this range (0.4–0.7 μm) is visible, while that in the infrared region (0.7–1.4 μm) is not visible.

Any damage that is restricted to the surface layer (up to 50 μm depth) of the cornea tends to heal in about 48 hours since the tissue in this region regenerates itself rather easily. However, if the deeper layers of the cornea are damaged, these are likely to result in scarring and the cornea

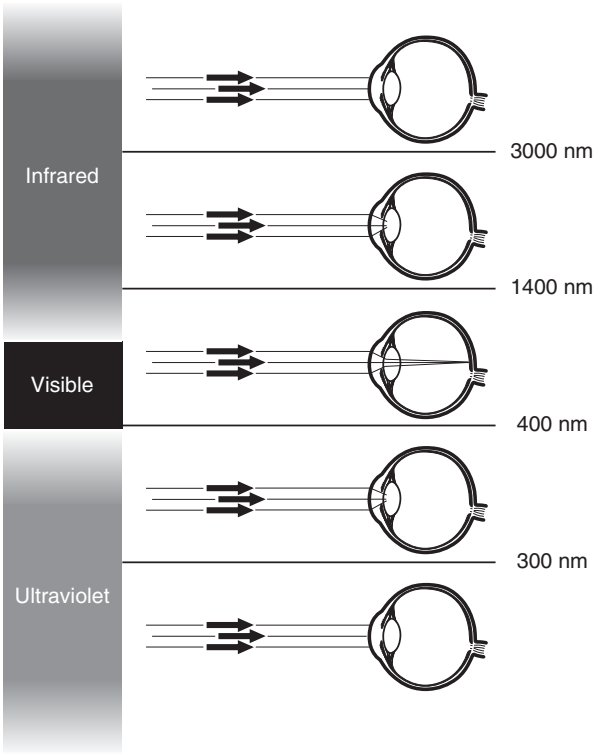


Figure 17.3 Absorption characteristics of the different components of the eye. Source: From Henderson (1997).

becoming permanently opaque. Damage to the lens may cause cataract (the lens becoming opaque). Photochemical damage may result in changes to night and color vision if the retina is subjected to long-term exposure at radiation levels that may not cause immediate injury. More severe damage to the eye includes the formation of a blind spot, distortion of vision, loss of high-resolution capability in the eye, and partial or total blindness.

Skin Hazards The extent to which different wavelengths of radiation penetrate the human skin is illustrated schematically in Figure 17.4. It is evident from this figure that the radiation in the visible wavelength regime results in the highest penetration. Ultraviolet and far-infrared radiation are mainly absorbed in the outer protective layer of dead cells. Damage to the skin includes sun burn and tanning (due to photochemical reaction resulting from ultraviolet radiation), inflammation due to dilation of blood capillaries and a localized laser burn (due to a photothermal action resulting from a high intense beam), premature aging and skin cancer (from prolonged exposure), and tissue cutting due to ablation.

17.1.2 Nonbeam Hazards

In addition to the hazards associated with the beam itself as outlined earlier, other forms of hazards are associated with the overall laser system and environment, for which appropriate precautions are necessary. Such hazards can be life threatening and include

1. Electrical hazards.

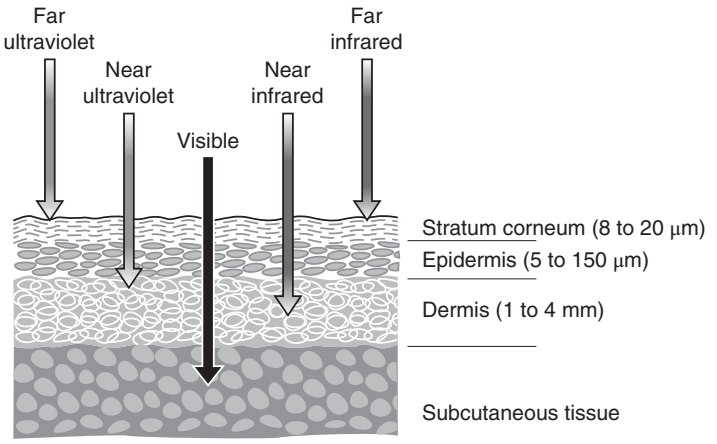


Figure 17.4 Extent to which different wavelengths of radiation penetrate the human skin. Source: From Henderson (1997).

2. Chemical hazards.
3. Environmental hazards.
4. Fire hazards.
5. Explosion hazards and compressed gases.
6. Other hazards.

We shall elaborate on some of these hazards, and the reader is referred to the American National Standards Institute (ANSI) Standards Z136.1-2007 or the International Electrotechnical Commission (IEC) Standards for further details.

17.1.2.1 Electrical Hazards A number of lasers need very high voltages for exciting the active medium, and this is especially the case for the high-power lasers. Electric shock can result from contact with voltages exceeding 50 volts. Such contacts usually occur when equipment protective covers are removed, for example during laser setup, installation, maintenance, service, etc. Capacitor banks that can store significant electric charge are often used in pulsed lasers. Deaths have occurred in connection with laser systems as a result of electric shock. The major sources of problems are as follows:

1. Uncovered electrical terminals.
2. Improperly insulated electrical terminals.
3. Improperly displayed power on warning lights.
4. Improperly grounded equipment.

Other problem sources are inadequate first-aid training, nonadherence to standards, and wires or cables on floor, creating slip hazards.

17.1.2.2 Chemical Hazards Some materials used in connection with lasers may be toxic and carcinogenic. These include some laser gases, especially those used with some excimer lasers, and some laser dyes.

17.1.2.3 Environmental Hazards Interaction of some lasers (especially the high-power ones) with matter may generate air contaminants such as fumes, dust, and vapors which may be toxic and/or noxious. The materials may be plastics, composites, metals, or tissue. Certain materials used for laser windows may also generate hazardous substances when a threshold beam intensity is exceeded. The impact of such contaminants can be significantly reduced by using

1. Proper exhaust ventilation.
2. Respiratory protection.
3. Process isolation.

For exhaust ventilation, it is preferable to use hoods whenever possible since they are more effective. Respiratory protection is normally used as a temporary measure. Process isolation involves isolating the system using physical barriers or using remote control.

17.1.2.4 Fire Hazards This is especially important for Class 4 lasers (Section 17.2). Accidental exposure of combustible material to high-power beams (exceeding 0.5 W) can cause ignition (author's personal experience in the laboratory), with the risk of starting a fire. Care should therefore be taken to avoid placing combustible materials in the beam path. For enclosure of the work area, flame-retardant materials should be used, whenever possible.

17.1.2.5 Explosion Hazards and Compressed Gases Some components of the laser system can be explosive if they eventually disintegrate. These include high-pressure arc lamps, capacitor banks, and some chemicals that may be used. In addition, proper precautions must be taken to ensure that the compressed gases used in a number of laser systems do not pose a health risk. Gas tanks must be properly secured, and not left free-standing. Hazardous gases must be appropriately labeled.

17.1.2.6 Other Hazards Other hazards associated with laser systems include mechanical hazards involving rotating machinery, and the use of robots, noise, ergonomics issues associated with handling workpieces and other laser components, and generation of other forms of radiation by system components.

In the next section, we outline the various classes of lasers. This, to a large extent, determines the precautions that need to be taken to avoid injury in using lasers.

17.2 LASER CLASSIFICATION

The classification of a laser is based primarily on the ability of the main or reflected beam to cause injury to the eye or skin, and that depends on the level of accessible emission for that laser. For a laser in a given class, the output of that laser must not exceed the *accessible emission limit* (*AEL*) for that class. The AEL is the maximum accessible emission level permitted within a particular class. In this section, we provide a brief outline of the major classes of lasers. The interested reader is referred to the ANSI Z136.1-2007 document on the "Safe Use of Lasers," for further details.

There are four major laser classes:

Class 1: The radiation levels produced by Class 1 lasers are not considered high enough to cause any form of injury either during operation or maintenance. Thus, no special precautions are necessary. An example is the laser in an audio compact disk player.

Class 2: These are low-power lasers and are further subdivided into 2 and 2M. The radiation from Class 2 lasers in general fall within the visible spectrum. Since the radiation is visible and of low power, natural reflexes such as blinking (with a response time of about 0.2 seconds) are considered adequate in protecting the eye against injury. However, damage may occur if one deliberately stares into the beam.

Class 3: These are medium-power lasers and are also further subdivided into 3R and 3B. This type of laser is not a skin hazard, but definitely an eye hazard. Direct and specularly reflected Class 3 beams can cause eye injury in a time period shorter than what it takes for the eye to blink, even though the diffusely reflected beam is not normally hazardous. It also does not pose a fire hazard.

Class 4: With the high-power inherent in this class of lasers, they can be a hazard to the eye and skin, as well as pose a fire hazard. There is not upper power limit to this class of lasers.

17.3 PREVENTING LASER ACCIDENTS

In this section, we outline some of the major steps that need to be taken to prevent or at least minimize the occurrence of laser accidents. A more detailed description of control measures and precautions that need to be taken are provided by the ANSI standard. The major issues that we shall touch on are as follows:

1. Laser safety officer (LSO).
2. Engineering controls.
3. Administrative and procedural controls.
4. Protective equipment.
5. Warning signs and labels.

17.3.1 Laser Safety Officer (LSO)

A LSO is someone who is given the responsibility and authority for ensuring safety in connection with the laser system. The appointment of an LSO is especially important for Class 3 and Class 4 laser installations. The LSO's responsibilities include all the control measures that are necessary for ensuring safe use of the laser as specified in the ANSI Standard Z136.1-2007, and summarized in Sections 17.3.2–17.3.5.

17.3.2 Engineering Controls

Engineering controls that are necessary for ensuring the safe use of lasers include the provision of protective housings, interlocks, key control, beam path enclosure, remote interlock connector, emergency stop button, laser activation warning systems, area control, equipment labels, and area sign posting.

Protective housing is normally provided by the equipment manufacturer. However, in situations where it may be necessary to use the laser without protective housing, such as in research and development, it will be necessary to provide such controls as access restriction and other safety measures. It is also necessary to provide protective housings with interlock systems which are activated to interrupt the beam, for example, when the protective housing is opened during operation or maintenance. Enclosure of the laser equipment and beam path is one of the most important steps for minimizing hazards.

The provision of a master switch with a key or coded access is necessary for Class 3B or 4 lasers, and in addition, a permanently attached emergency stop button should be provided. Laser activation warning systems such as an alarm or warning light that indicate activation of high voltage and radiation emission are also necessary. It is preferable to have such warnings activated before the emission of laser radiation.

It is also necessary to enclose the beam path. Where this is not completely possible, the LSO will establish a nominal hazard zone (NHZ), unless this is already provided by the laser manufacturer. A *NHZ* describes the space within which the level of the direct, reflected, or scattered radiation during operation exceeds the applicable maximum permissible exposure (MPE). A MPE, as shown in Tables 17.1 and 17.2, is the level of radiation to which a person may be exposed without hazardous effect or adverse biological changes in the eye or skin. Under such circumstances, the work area should be designated as a laser-controlled area with restricted access which is limited only to necessary personnel. A remote interlock connector which terminates laser emission when opened is then necessary for Class 3B and 4 lasers. Such a system is illustrated in Figure 17.5. It may also be designed with an interlock-override capability so that appropriate personnel can enter and leave the facility without necessarily interrupting the laser operation.

17.3.3 Administrative and Procedural Controls

Administrative and procedural controls specify rules and/or work practices that implement engineering controls. These include standard operating procedures which are maintained with the laser equipment for reference by the operator, maintenance, or service personnel. They also include specifying limitations on output radiation emission, education and training, specification of authorized personnel (operators, maintenance, and service personnel), alignment procedures, specification of protective equipment, and control of spectators.

17.3.4 Protective Equipment

The most important piece of protective equipment necessary for minimizing laser hazards is protective eyewear. Others include windows, barriers, curtains, and skin protective devices.

17.3.4.1 Protective Eyewear Eye protection is essential when personnel are either working with or are in the NHZ of Class 3B or 4 lasers that are operating. Its main purpose is to reduce any possible exposure to safe levels (no more than the MPE), while at the same time permitting enough radiation to be transmitted across the visible spectrum such that visual performance is not unduly degraded. Thus, the transmittance of the filter at the laser wavelength should not exceed $\tau_{\lambda\max}$, which is given by:

$$\tau_{\lambda_{\max}} = \frac{\text{MPE}}{\text{maximum possible exposure}} \tag{17.1}$$

Protective devices that can be used include goggles, face shields, and spectacles or prescription eyewear using special filter materials or reflective coatings. Such protective eyewear should withstand either direct or diffusely scattered beams. Important factors that need to be considered in selecting protective eyewear are as follows:

1. Wavelength(s) of laser output.
2. Potential for multiwavelength operation.
3. Radiant exposure or intensity level.
4. Potential exposure time.
5. MPE.
6. Optical density (OD) requirement of eyewear filter at laser output wavelength.

Table 17.1 Maximum permissible exposure for the eye.

Wavelength (μm)	Exposure duration, t (s)	MPE (J/cm^2)	MPE (W/cm^2)	Notes
Ultraviolet				In the dual-limit wavelength region (0.180 to 0.400 μm), the lower MPE considering photochemical and thermal effects must be chosen
	Dual limits for λ between 0.180 and 0.400 μm			
Thermal				
0.180 to 0.400	10^{-9} to 10	$0.56\ t^{0.25}$		
Photochemical				
0.180 to 0.302	10^{-9} to 3×10^{4}	3×10^{-3}		
0.302 to 0.315	10^{-9} to 3×10^{4}	$10^{200(\lambda-0.295)}\times10^{-4}$		
0.315 to 0.400	10 to 3×10^{4}	1.0		
Visible				
0.400 to 0.700	10^{-13} to 10^{-11}	1.5×10^{-8}		In the wavelength region (0.400 to 0.500 μm), T_1 determines whether the photochemical or thermal MPE is lower
0.400 to 0.700	10^{-11} to 10^{-9}	$2.7\ t^{0.75}$		
0.400 to 0.700	10^{-9} to 18×10^{-6}	5.0×10^{-7}		
0.400 to 0.700	18×10^{-6} to 10	$1.8\ t^{0.75}\times10^{-3}$		
0.500 to 0.700	10 to 3×10^{4}		1×10^{-3}	
Thermal				
0.450 to 0.500	10 to T_1		1×10^{-3}	
Photochemical				
0.400 to 0.450	10 to 100	1×10^{-2}		
0.450 to 0.500	T_1 to 100	$C_B\times10^{-2}$		
0.400 to 0.500	100 to 3×10^{4}		$C_B\times10^{-4}$	
Near infrared				
0.700 to 1.050	10^{-13} to 10^{-11}	$1.5\ C_A\times10^{-8}$		
0.700 to 1.050	10^{-11} to 10^{-9}	$2.7\ C_A\ t^{0.75}$		
0.700 to 1.050	10^{-9} to 18×10^{-6}	$5.0\ C_A\times10^{-7}$		
0.700 to 1.050	18×10^{-6} to 10	$1.8\ C_A\ t^{0.75}\times10^{-3}$		
0.700 to 1.050	10 to 3×10^{4}		$C_A\times10^{-3}$	
1.050 to 1.400	10^{-13} to 10^{-11}	$1.5\ C_C\times10^{-7}$		The wavelength region λ_1 to λ_2 means $\lambda_1 \le \lambda < \lambda_2$, e.g. 0.180 to 0.302 μm means $0.180 \le \lambda < 0.302$ μm
1.050 to 1.400	10^{-13} to 10^{-11}	$27.0\ C_C\ t^{0.75}$		
1.050 to 1.400	10^{-9} to 50×10^{-6}	$5.0\ C_C\times10^{-6}$		
1.050 to 1.400	50×10^{-6} to 10	$9.0\ C_C\ t^{0.75}\times10^{-3}$		
1.050 to 1.400	10 to 3×10^{4}		$5.0\ C_C\times10^{-3}$	
Far infrared				Note: The MPEs must be in the same units.
1.400 to 1.500	10^{-9} to 10^{-3}	0.1		
1.400 to 1.500	10^{-3} to 10	$0.56\ t^{0.25}$		
1.400 to 1.500	10 to 3×10^{4}		0.1	
1.500 to 1.800	10^{-9} to 10	1.0		
1.500 to 1.800	10 to 3×10^{4}		0.1	
1.800 to 2.600	10^{-9} to 10^{-3}	0.1		
1.800 to 2.600	10^{-3} to 10	$0.56\ t^{0.25}$		
1.800 to 2.600	10 to 3×10^{4}		0.1	
2.600 to 1000	10^{-9} to 10^{-7}			
2.600 to 1000	10^{-7} to 10	$0.56\ t^{0.25}$		
2.600 to 1000	10 to 3×10^{4}		0.1	

Source: From American National Standard for Safe Use of Lasers ANSI Z136.1-2007. Copyright 2007, Laser Institute of America. www.laserinstitute.org. All rights reserved.

Table 17.2 Maximum permissible exposure for the skin.

Wavelength (μm)	Exposure duration, t (s)	MPE	
		(J/cm^2) except as noted	(W/cm^2) except as noted
Ultraviolet			
	Dual Limits for λ between 0.180 to 0.400 μm		
Thermal			
0.180 to 0.400	10^{-9} to 10	0.56 $t^{0.25}$	
Photochemical			
0.180 to 0.302	10^{-9} to 3×10^4	3×10^{-3}	
0.302 to 0.315	10^{-9} to 3×10^4	$10^{200(\lambda - 0.295)} \times 10^{-4}$	
0.315 to 0.400	10 to 10^3	1.0	
0.315 to 0.400	10^3 to 3×10^4		1×10^{-3}
Visible and near infrared			
0.400 to 1.400	10^{-9} to 10^{-7}	2 $C_A \times 10^{-2}$	
0.400 to 1.400	10^{-7} to 10	1.1 C_A $t^{0.25}$	
0.400 to 1.400	10 to 3×10^4		0.2 C_A
Far infrared			
1.400 to 1.500	10^{-9} to 10^{-3}	0.1	
1.400 to 1.500	10^{-3} to 10	0.56 $t^{0.25}$	
1.400 to 1.500	10 to 3×10^4		0.1
1.500 to 1.800	10^{-9} to 10	1.0	
1.500 to 1.800	10 to 3×10^4		0.1
1.800 to 2.600	10^{-9} to 10^{-3}	0.1	
1.800 to 2.600	10^{-3} to 10	0.56 $t^{0.25}$	
1.800 to 2.600	10 to 3×10^4		0.1
2.600 to 1000	10^{-9} to 10^{-7}	1×10^{-2}	
2.600 to 1000	10^{-7} to 10	0.56 $t^{0.25}$	
2.600 to 1000	10 to 3×10^4		0.1

Source: From American National Standard for Safe Use of Lasers ANSI Z136.1-2007. Copyright 2007, Laser Institute of America. www.laserinstitute.org. All rights reserved.

The optical density, OD_λ, is the factor by which incident radiation at a specific wavelength would be attenuated and is given by:

$$OD_\lambda = \log_{10} \frac{I_p}{MPE} = -\log_{10} \tau_\lambda \tag{17.2}$$

where I_p is the highest level of intensity (power per unit area) to which the eye could be exposed and is given by the beam power over the limiting aperture (which is 7 mm for the 0.4 to 1.4 μm region) (W/cm^2), MPE is the maximum permissible exposure (W/cm^2), and τ_λ is the transmittance of the filter at a given wavelength.

I_p and MPE have to be in the same units. The value of OD_λ that is calculated from Eq. (17.2) is rounded up to the next whole number. For a given potential laser output, the MPE for that wavelength has to be determined. Then, from Eq. (17.2), the OD which will indicate the type of protective eyewear to be used can be calculated. For example, if the laser exposure is 10^5 times greater than the MPE, then an eyewear with an OD of at least 5 would be needed.

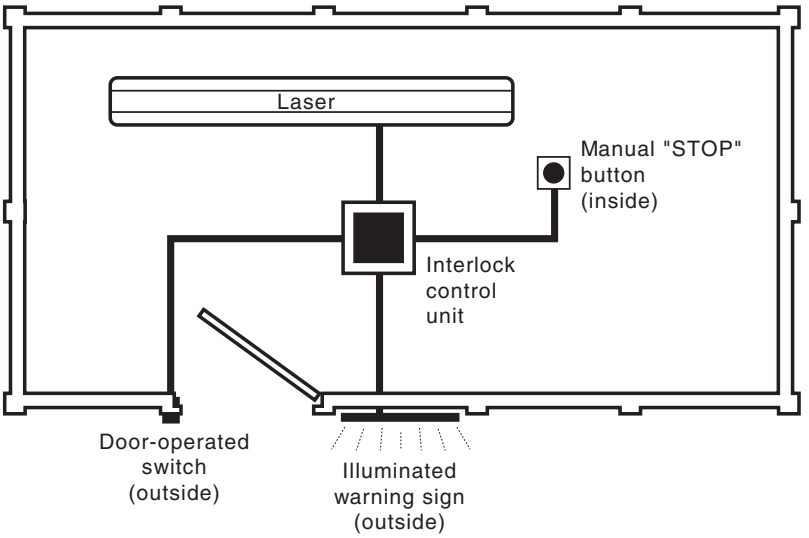


Figure 17.5 Illustration of interlocked access to a laser hazard area. Source: From Henderson (1997).

Table 17.3 Relationship between the optical density, transmittance, and level of protection.

Transmittance	Level of protection MPE	OD_λ
1	1	0
10^{-1}	10	1
10^{-2}	10^2	2
10^{-3}	10^3	3
10^{-4}	10^4	4
10^{-5}	10^5	5
10^{-6}	10^6	6
10^{-7}	10^7	7

Table 17.3 shows the relationship between the OD, transmittance, etc., at a given wavelength and indicates, for a calculated OD for that wavelength, the percentage of the beam that will be transmitted, or attenuated, or the level of protection to be expected.

For ease of reference, ODs necessary for various laser outputs are summarized in Table 17.4. For completeness, the MPE for both eye and skin exposure are provided in Tables 17.1 and 17.2, respectively. It must be noted that the MPE value used depends on the duration of exposure to the beam.

17.3.4.2 Other Protective Equipment Other protective equipment include laser protective windows, barriers and curtains, and skin protection clothing such as gloves or even "sunscreen" creams.

Example 17.1 As a Laser Systems Engineer with XYZ Company, you are in charge of installing a new 3 kW Nd:YAG laser system in the plant. Calculate an optical density for the eyewear to be used for protection against long-term exposure.

Table 17.4 Optical densities for various laser outputs.

Q-switched laser (10^{-9}–10^{-2} s)		Non-Q-switched lasers (0.4×10^{-3}–10^{-2} s)		Continuous-wave lasers momentary (0.25–10 s)		Continuous-wave lasers long-term staring (<1 h)		Attenuation	
Maximum output energy (J)	Max beam radiant exposure (J/cm^2)	Max laser output energy (J)	Max beam radiant exposure (J/cm^2)	Max power output (W)	Max beam irradiance (W/cm^2)	Max power output (W)	Max beam irradiance (W/cm^2)	Attenuation factor	OD
10	20	100	200	10^{5a}	2×10^{5a}	100^a	200^a	10^8	8
1	2	10	20	10^{4a}	2×10^{4a}	10^a	20^a	10^7	7
10^{-1}	2×10^{-1}	1	2	10^{3a}	2×10^{3a}	1	2	10^6	6
10^{-2}	2×10^{-2}	10^{-1}	2×10^{-1}	100^a	200^a	10^{-1}	2×10^{-1}	10^5	5
10^{-3}	2×10^{-3}	10^{-2}	2×10^{-2}	10	20	10^{-2}	2×10^{-2}	10^4	4
10^{-4}	2×10^{-4}	10^{-3}	2×10^{-3}	1	2	10^{-3}	2×10^{-3}	10^3	3
10^{-5}	2×10^{-5}	10^{-4}	2×10^{-4}	10^{-1}	2×10^{-1}	10^{-4}	2×10^{-4}	10^2	2
10^{-6}	2×10^{-6}	10^{-5}	2×10^{-5}	10^{-2}	2×10^{-2}	10^{-5}	2×10^{-5}	10	1

[a]Not recommended as a control procedure at these levels. These levels of power could damage or destroy the attenuating material used in the eye protection. The skin also needs protection at these levels.
Use of this table may result in optical densities (ODs) greater than necessary.
Source: From American National Standard for Safe Use of Lasers ANSI Z136.1-2007. Copyright 2007, Laser Institute of America. www.laserinstitute.org. All rights reserved.

Solution:
Wavelength of the Nd:YAG laser is 1.06 μm.
From Table 17.1, the MPE for long-term exposure for this wavelength is 5.0 $C_c\times10^{-3}$ W/cm^2.
From Table 17.5 (Table 6 of ANSI Standards Z136.1-2007), $C_c = 1.0$ for the 1.06 μm wavelength.
Thus, MPE $= 5.0\times10^{-3}$ W/cm^2.

Table 17.5 Correction factors.

Parameters/correction factors	Wavelength (μm)
$C_A = 1.0$	0.400 to 0.700
$C_A = 10^{2(\lambda-0.700)}$	0.700 to 1.050
$C_A = 5.0$	1.050 to 1.400
$C_B = 1.0$	0.400 to 0.450
$C_B = 10^{20(\lambda-0.450)}$	0.450 to 0.600
$C_C = 1.0$	1.050 to 1.150
$C_C = 10^{18(\lambda-1.150)}$	1.150 to 1.200
$C_C = 8.0$	1.200 to 1.400
$T_l = 10\times10^{20(\lambda-0.450)a}$	0.450 to 0.500

[a]$T_l = 10$ s for $\lambda = 0.450$ μm, and $T_l = 100$ s for $\lambda = 0.500$ μm
Note 1: Wavelengths must be expressed in micrometers for calculations.
Note 2: The wavelength region λ_1 to λ_2 means $\lambda_1 \le \lambda < \lambda_2$, 0.550 to 0.700 μm means $0.550 \le \lambda < 0.700$ μm.
Source: From American National Standard for Safe Use of Lasers ANSI Z136.1-2007. Copyright 2007, Laser Institute of America. www.laserinstitute.org. All rights reserved.

Using a limiting aperture of 7 mm = 0.7 cm, the highest level of intensity to which the eye could be exposed, I_p, is given by:

$$I_p = \frac{3000}{\pi 0.7^2/4} = 7794 \text{ W/cm}^2$$

Therefore,

$$\text{OD}_\lambda = \log_{10}\frac{I_p}{\text{MPE}} = \log_{10}\frac{7794}{5.0 \times 10^{-3}} = -\log_{10}(1.56 \times 10^6) = 6.19$$

which gives a final rounded-up OD of 7. This compares with a value of 8 to be read directly from Table 17.4. Correction factors used in various tables are listed in Table 17.5.

17.3.5 Warning Signs and Labels

Another effective means of preventing accidents involves the use of warning signs and labels on the lasers. These inform anyone that uses the laser about the type of hazard associated with it. Such signs may also be placed at the entrance to the NHZ of the laser. The design of signs in terms of dimensions, color, letter size, and so on are specified in the American National Standard Specification for Accident Prevention Signs, ANSI Z535. The ANSI standards for signs used to denote Class 2, 3, and 4 lasers are shown in Figure 17.6. For Class 2 lasers, the word "Caution" is used with all signs and labels, while for 3B and 4 lasers, the word "Danger" is used. The International Standard is based on the symbol as shown in Figure 17.7 (IEC Publication 825). Additional wording that is required, depending on the type of laser is as follows:

Class 2: "Laser Radiation - Do Not Stare into Beam."

Class 2M and Some 3R: "Laser Radiation - Do Not Stare into Beam or View Directly with Optical Instruments.

Class 3R (others): "Laser Radiation - Avoid Direct Eye Exposure."

Class 3B: "Laser Radiation - Avoid Direct Exposure to Beam."

Class 4: "Laser Radiation - Avoid Eye or Skin Exposure to Direct or Scattered Radiation."

Other pertinent information to be provided on the sign includes

Position 1: Invisible Laser Radiation, Knock Before Entering, Do Not Enter When Light is On, Restricted Area, etc.

Position 2: Type of laser, emitted wavelength, pulse duration, maximum output.

Position 3: Class of the laser.

17.4 SUMMARY

Lasers need to be used with caution for safety, since improper usage can cause serious injury. Injuries associated directly with the laser beam generally occur with respect to the human skin and eye. The damage that is sustained depends on the power density, wavelength, the irradiated area and its absorption characteristics, and the exposure duration.

Damage to human tissue may result from photothermal, photochemical, or photoacoustic effects. With regard to the eye, focusing of a collimated beam on the retina results in magnification of the light intensity by a factor of about 10^5. Thus relatively low-intensity radiation can cause significant

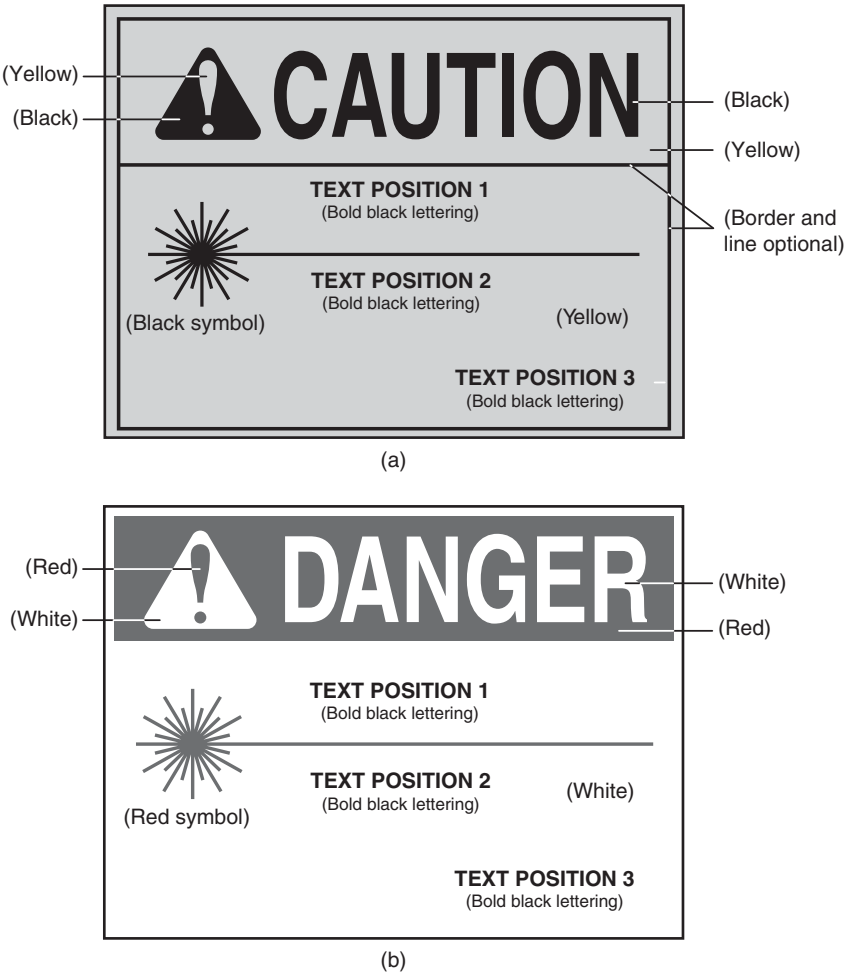


Figure 17.6 ANSI warning signs for lasers. (a) Class 2 and 2M lasers. (b) Class 3R, 3B, and 4 lasers. Source: From American National Standard for Safe Use of Lasers ANSI Z136.1-2007. Copyright 2007, Laser Institute of America. www.laserinstitute.org. All rights reserved.

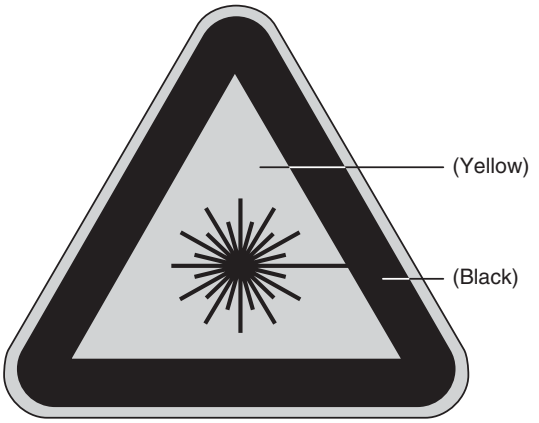


Figure 17.7 International warning signs for lasers.

damage to the retina. This is especially so for radiation in the visible and near-infrared regions. Damage to the surface layer of the cornea tends to heal in about 48 hours. However, damage to deeper layers of the cornea can result in scarring and the cornea becoming permanently opaque. Damage to the lens may cause cataract. More severe damage to the eye includes the formation of a blind spot, distortion of vision, loss of high-resolution capability in the eye, and partial blindness. With regard to the skin, radiation in the visible wavelength regime results in the highest penetration. Ultraviolet and far-infrared radiation are mainly absorbed in the outer protective layer of dead cells. Damage to the skin includes sun burn and tanning, inflammation due to dilation of blood capillaries, premature aging, and skin cancer.

There are other hazards that do not directly involve the beam, and these are associated with electrical, chemical, environmental, fire, explosion, and compressed gas sources.

There are four major laser classes. For each class, the laser output must not exceed the AEL for that class. Details are provided in the ANSI Standard Z136.1-2007. The four classes are as follows:

Class 1: Radiation levels for this class of lasers are not high enough to cause any form of injury, for example, the laser in a compact disk player.

Class 2: These are low-power lasers, fall within the visible spectrum, and thus natural reflexes such as blinking are adequate in protecting the eye against injury. However, damage may occur if one deliberately stares into the beam.

Class 3: These are medium-power lasers, and even though not a skin hazard, are definitely an eye hazard. They also do not pose a fire hazard.

Class 4: With the high-power inherent in this class of lasers, they can be a hazard to the eye and skin, as well as pose a fire hazard. There is no upper power limit to this class of lasers.

To ensure that appropriate steps are taken to minimize laser accidents, it is necessary to have

1. *A Laser Safety Officer*: who ensures safety in connection with the laser system.
2. *Engineering Controls*: which include the provision of protective housings, interlocks, beam path enclosure, emergency stop button, laser activation warning systems, area sign posting, etc.
3. *Administrative and Procedural Controls*: which specify rules and/or work practices that implement engineering controls.
4. *Protective Equipment*: which include protective eyewear.
5. *Warning Signs and Labels*: which inform anyone that uses the laser about the type of hazard associated with it.

APPENDIX 17.A

List of symbols used in the chapter

Symbol	Parameter	Units
AEL	Accessible emission limit	W
C_A, C_B, C_C	Correction factors	—
I_p	Highest irradiance level	W/cm^2
MPE	Maximum permissible exposure	W/cm^2
NHZ	Nominal hazard zone	—
OD_λ	Optical density	—
τ_λ	Filter transmittance at laser wavelength	—

PROBLEM

17.1. Show that the eye magnifies the intensity of a collimated beam by about 10^5 times.

BIBLIOGRAPHY

Henderson, A.R. (1997). *A Guide to Laser Safety*. London: Chapman and Hall.

Laser Institute of America. (2007). *American National Standard for Safe Use of Lasers*, ANSI Z136.1-2007. Orlando, FL: Laser Institute of America.

Matthews, L. and Garcia, G. (1995). *Laser and Eye Safety in the Laboratory*. New York: IEEE Press, The Institute of Electrical and Electronic Engineers, Inc.

Winburn, D.C. (1990). *Practical Laser Safety*, 2e. New York: Marcel Dekker, Inc.

INDEX

A posteriori probability 535, 536, 550, 552
A priori probability 533, 534, 535, 538, 550, 552
ablation 69, 264, 275, 279, 280, 312, 318–326, 326, 327, 346, 347, 353, 412, 420, 422, 423, 429, 486, 488, 490–492, 494, 560
absorption
  Fresnel 289, 348, 349, 382
  inverse bremsstrahlung 289, 290, 320, 348, 382, 422
  stimulated 4, 5, 43
absorptivity 89, 269, 270, 271, 276, 299, 300, 305, 308, 326, 344, 381, 397, 399, 427, 503
acoustic emission 197, 497, 504, 507, 509, 510, 534, 535
activation energy 180, 181, 207, 402, 403, 426, 432, 433
active medium 9–11, 17, 20–23, 31, 33, 43, 47, 51, 55–58, 60–62, 64, 80, 81, 84, 85, 90, 396, 561
active region 74–77
additive manufacturing 263, 453, 454, 460–464, 468, 470, 472, 473, 475, 477–481
aliasing 527, 531–532, 550
angle of acceptance 106, 107, 113, 126
angle of incidence 97, 98, 108, 125, 283, 290, 291, 397, 399, 400, 412,
aperture 18, 36, 37, 43–44, 52, 76, 79, 85, 104, 117–119, 122, 123, 248, 504, 566, 569
arc-augmented laser welding 335, 367
argon (Ar^+) ion laser 90, 467, 483, 490
attenuation 105, 114, 116, 470, 499, 502, 517, 568
audible sound 197, 504, 509, 510
austenite 180, 192–195, 199, 200, 227, 237, 238, 392, 393, 400, 402–407, 410, 411, 413, 430–433
avalanche ionization 319, 320
axial mode 14, 15, 49, 51, 60. *See also* Longitudinal
Bayes rule 534, 535
beam
  alignment 32, 263, 271, 272, 275
  defocusing 393, 394, 396
  delivery 95–130, 263, 270–275, 281, 326, 335, 370, 394, 437, 509
  divergence 18, 36, 37, 49, 52, 76, 90, 267, 272, 307, 497
  dump 498, 499, 522, 523
  expander 101–102, 104, 129, 277, 278
  focusing 38, 42, 170, 361
  intensity 8, 9, 17, 34, 75, 87, 88, 247, 276, 279, 290, 291, 302, 310, 319, 326, 327, 340, 363, 424, 465, 488, 491, 492, 501, 503, 562
  mode 14, 116, 263, 264, 266, 275, 282, 283, 319, 336–337, 382, 393–394, 427, 497, 499, 503
  power 43, 102, 151, 263, 265, 275, 282, 283, 289, 313, 326, 327, 335, 336, 339, 367, 382, 393, 438, 490, 497–499, 501, 503, 507, 557, 566
  quality 24, 78–82, 85, 88–90, 116, 120, 263, 267–269, 275, 310, 311, 327, 385
  size 19, 20, 43, 48, 52, 82, 101, 102, 104, 116, 122, 263, 264, 266, 272, 277, 318, 319, 342, 369, 381, 393, 428, 465, 473, 476, 497, 499, 503, 504, 522
  waist 20, 49, 122, 266, 269, 278, 385
beam-shaping 393, 396, 397
Beer-Lambert law 8, 380
bessel function 141, 142, 167
Bezier curve 456–458, 481
birefringence 95, 96
blackbody radiation 7, 264, 275, 512
Boltzmann's constant 4, 24, 28, 42, 70, 179, 314, 316
Boltzmann's law 4, 5, 24, 43, 71

Principles of Laser Materials Processing: Developments and Applications, Second Edition. Elijah Kannatey-Asibu, Jr.

© 2023 John Wiley & Sons, Inc. Published 2023 by John Wiley & Sons, Inc.

Companion Website: www.wiley.com/go/Kannatey-Asibu/PrinicplesLaserMaterialsProcessing

boundary representation 459–460
Bragg's law 245, 246
brewster angle 59, 65, 95, 96, 97, 101, 125
brightness 35, 42, 85, 88, 518, 520
broadening
  collision 26, 27, 29, 44, 50, 57, 65
  Doppler 28, 29, 44, 50, 57, 60, 65
  homogeneous 26–28, 44
  inhomogeneous 26, 28–29, 44, 57
buckling mechanism 440–443, 448
butt joint 342–344, 375, 382

CAD (computer-aided design) 453–462, 470, 478, 479, 481, 483
carbon dioxide laser, *see* laser – carbon dioxide
cavity
  lifetime 30, 46
  losses 24, 32, 35, 76
  mode(s) 12, 56
cladding 67, 84–88, 106–109, 112, 113, 126–130, 135, 161, 171, 175, 179, 207, 216, 264, 391, 397, 414–419, 429, 433–435,
classification 196, 528, 534, 537, 539, 562
coating 75, 102, 103, 124, 159, 198, 269, 300, 323, 325, 326, 355, 371, 373, 379, 380, 383, 391, 392, 397–399, 414, 415, 419, 420, 422, 423, 429, 434, 436, 438, 439, 463, 478, 481, 564
coherence
  length 41, 52
  spatial 38, 40, 52
  temporal 38–40, 52
  time 39, 41, 42, 46, 52
conduction band 69–75, 90, 91, 114, 320
conduction mode welding 79, 90, 138, 158, 341, 345, 348, 368, 369, 382
confocal resonator 18–20, 48, 52
constructive solid geometry (CSG) 459
continuity equation 156, 293, 296
continuous wave (CW) 22, 56–61, 76–78, 82, 84, 263, 283, 335, 488, 497
cooling rate 136, 148–150, 161, 169–171, 184, 188, 192, 193, 197, 200, 203, 220, 362–364, 384–386, 400, 411, 497
cornea 558, 559, 571
cracking
  cold 197, 200, 203, 208, 390
  hot 188, 195, 197–200, 203, 355–357, 373
critical angle 106, 127
cut quality 283–285, 299, 327, 332, 509, 510
cutting
  clean 280
  fusion 279, 280, 288, 326, 331, 333
  sublimation 279, 280, 326
cutting front 290, 292–294, 327, 328, 332, 509, 511

data reduction 527, 532, 535, 550,
dentistry 486, 488, 489, 494
depth of focus 265, 275–277, 279, 311, 330, 336
dermatology 119, 486, 488, 489, 494
dielectric 10, 12, 13, 85, 105, 110, 319, 320, 322, 327, 411, 420–423, 429, 432, 505
diffraction
  neutron 227, 240, 247, 248, 252, 259
  x-ray 227, 240, 245–247, 249, 252, 257, 258
diffraction-limited 36, 37, 52, 87, 88, 268, 269, 276
diffraction losses 20, 394, 396
dilution 414, 417–419
diode (semiconductor) laser, *see* laser – diode (semiconductor)
direct metal deposition 391, 414, 471, 473
discriminant function 536, 537, 539, 540, 550–555
disk laser 55, 80–82, 89, 90, 93
dispersion 33, 105, 108–113, 116, 126, 127, 129, 217, 326, 398
dissimilar materials 120, 335, 355, 360, 381, 386
dissimilar metals 358, 371, 381, 383, 388, 514
distortion 104, 109, 111, 112, 113, 126, 129, 135, 191, 194, 227–252, 305, 310, 324, 325, 346, 360, 412, 419, 478, 518, 530, 547, 560, 571
divergence 18, 35–37, 44, 49, 52, 75, 76, 79, 87, 90, 263, 267, 272, 288, 307, 497, 503, 544, 545
doping 71, 84, 85, 159
Doppler 15, 28, 29, 44, 50, 57, 60, 65
Doppler broadening 28, 28, 29, 44, 50, 57, 60, 65
dross 267, 280, 283, 284, 285, 286, 296, 297–301, 318, 321, 327, 486, 487, 497, 504
ductility 148, 187, 197, 203, 408, 410
dye laser 55, 61, 489, 561

efficiency
  heat transfer 144, 147, 150, 171, 273, 276, 303, 329, 385
  laser (generator) 273, 276,
  process 263, 272–274, 283, 338, 339, 340, 344
  pumping 21, 22, 86
  transmission 273, 276
  welding 335, 344
Einstein coefficient 5, 6, 8
electric field 10, 23, 24, 30, 33, 34, 48, 66, 70, 74, 98, 113, 119, 122, 348, 505
electric vector 96, 98, 100, 110, 125, 399
electrical discharge 23, 66, 69, 305, 306, 307

electrical hazards 560, 561
electromagnetic spectrum 3, 4, 95–96, 501, 511, 516
electromagnetic wave 10, 12, 33, 38, 95, 98, 99, 100, 125, 395, 515
electron beam welding 274, 335, 344, 359, 360, 365
electronic energy level 61, 67, 68
electro-optic 32, 33, 35, 44, 103, 124
emissivity 512, 513, 516
emission
  spontaneous 5–9, 26, 37, 38, 44, 47, 50
  stimulated 3, 5, 6, 7, 9, 10, 43, 47, 73, 75
energy
  band 73
  gap 75
energy density 5–7, 47
energy levels
  electronic 61, 67, 68
  rotational 61
  translational 61
  vibrational 61
energy source 146, 311, 413, 447, 462
evaporation 288, 289, 295, 300, 314, 316, 318, 321, 322, 336, 345, 347, 417, 558
excimer laser 62, 67–69, 90, 93, 273, 318, 412, 491, 493, 561
excitation 11, 21, 23, 24, 26, 43, 57, 60, 64, 67, 71, 75, 88, 114
exposure 227, 252, 263, 400, 412, 426, 466, 467, 468, 482, 483, 558, 560, 562, 564–569
eye 3, 78, 88, 557, 558, 559, 560, 562–564, 565, 566, 567, 568, 569, 571

Fabry-Perot 14, 16, 18, 103
fast axial flow laser 65, 66
feature selection 532, 533, 546
femtosecond laser 12, 82–84, 311, 322, 323, 428, 470, 490, 491, 492
fermi energy level 70
ferrite 192, 193, 195, 199, 400, 402, 403, 404, 405, 413, 432
fiber laser 55, 84–90, 124, 273, 335, 336, 355, 356, 369, 379, 382, 383, 473
fiber optic 22, 23, 79, 82, 89, 90, 95, 103, 105–117, 126, 129, 272, 275, 360, 437, 520
focal length 43, 89, 102, 118, 122, 129, 264, 265, 275, 277, 278, 307, 310, 311, 330, 342, 369, 380, 383, 385, 394, 558
focal position 288, 342
four-level system 25–26, 44
Fourier transform 30, 527–528, 530, 546, 547, 550
frequency 3, 5, 6, 7, 10, 12, 13, 14, 16, 18, 19, 23, 24, 26, 28, 29, 30, 33–35, 37, 38, 43, 44, 47, 48, 50, 51, 52, 64, 73, 83, 85, 88, 95, 96, 98, 99, 107, 108, 125, 126, 128, 129, 213, 252, 269, 297, 310, 320, 324, 338, 367, 396, 412, 422, 429, 504, 505, 506, 507, 508, 509, 517, 524, 527–532, 546, 547, 548, 549, 550
fresnel absorption 289, 348, 349, 382
fused deposition modeling 454, 474–475
fusion cutting 279, 280, 288, 326
fusion welding 335, 381
fusion zone 143, 146, 147, 158, 161, 169, 171, 175, 176, 179, 180, 181, 184, 185, 187, 188, 191, 192, 195, 196, 197, 198, 200, 201, 202, 203, 209, 274, 413, 414, 415, 506

Gallium arsenide (GaAs) 73, 76, 77, 104, 265, 275, 501
gain 9, 10, 31, 50, 64, 76, 77, 82, 85, 87, 88, 89, 136, 506, 507, 543
gas jet 281, 287, 288, 289, 291, 293, 295, 299, 300, 307, 308, 310, 327, 509, 511
gas lasers
  argon ion 32, 33, 60, 90, 463, 467, 483, 489, 490
  carbon dioxide (*see* lasers – carbon dioxide)
  excimer 62, 67–69, 90, 93, 273, 318, 412, 491, 493, 561
  He-Ne (Helium-Neon) 57–60, 78, 90, 92, 129
Gaussian beam 37, 42, 49, 120, 121, 122, 151, 264, 267, 268, 269, 276, 277, 291, 327, 382, 394, 397, 398, 467, 482
Gaussian distribution 16, 29, 135, 138, 151, 161, 536
Gaussian mode 17, 53, 116, 120, 266, 276, 277, 393
glass 55, 56, 57, 69, 82, 85, 88, 89, 90, 98, 105, 106, 114, 115, 116, 129, 218, 238, 239, 269, 302, 319, 322, 411, 412, 420, 422, 423, 429, 468, 502
graded-index fiber 108–109, 116
grain size 188, 189, 190, 206, 393, 400, 406, 407, 411, 432, 433
ground state 4, 23, 56, 58, 60, 62, 63, 64, 67, 69, 86
group index 110, 128
group velocity 110, 111
growth 175, 176, 181, 182, 183, 185, 186, 187, 188, 191, 192, 194, 196, 202, 203, 209, 211, 212, 213, 214, 221, 222, 408, 411, 420, 442, 471, 486, 489

hardening 175, 188, 191, 192, 200, 203, 204, 240, 275, 325, 360, 391, 392, 399, 400, 407, 408, 412, 429, 444
hardness 193, 194, 200, 375, 391, 392, 393, 394, 400, 401, 407–411, 413, 414, 419, 420, 421, 422, 429, 497
hazard 302, 399, 557–564, 567, 569, 571
heat affected zone (HAZ) 143, 175, 188, 198, 204, 275, 279, 345, 360, 367, 373, 380, 396, 411, 412, 413, 419, 447, 488, 497, 506
heat flow 135, 136–156, 161, 162, 163, 169, 175, 181, 203, 214, 221, 222, 292, 295, 304, 314, 345, 350, 401, 446, 516
heat source 135–142, 143, 145, 146, 148, 151, 153, 161–164, 169, 170, 171, 194, 200, 201, 228, 229, 230, 238, 239, 263, 264, 274, 275, 284, 295, 304, 314, 339, 345, 348, 350, 360, 361, 362, 363, 364, 365, 371, 380, 381, 385, 386, 401, 410, 419, 440, 441, 443, 447, 448, 450, 488, 516, 517
heat transfer 144, 147, 150, 151, 163, 171, 239, 273, 276, 303, 329, 379, 385, 443
heat treatment 37, 38, 67, 79, 88, 116, 119, 137, 151, 161, 176, 193, 194, 195, 201, 227, 238, 249, 251, 263, 269, 270, 358, 360, 361, 363, 391–414, 426, 429, 432, 433, 437, 473
helium 58, 59, 340, 341, 342, 382, 393, 414
helium-neon (He-Ne) laser, *see* lasers – helium-neon
heterojunction laser 76–77
homogenization 400, 402, 404–406
homojunction laser 76
humping 359, 365, 366, 367, 382

inert gas 58, 60, 280, 284, 285, 286, 295, 298, 299, 301, 303, 327, 341, 414, 472
infrared 3, 15, 43, 56, 57, 59, 62, 76, 80, 84, 104, 115, 197, 271, 300, 394, 422, 469, 470, 481, 491, 501, 504, 510–516, 518, 522, 546, 558, 559, 560
inhomogeneous broadening 26, 28–29, 44, 57
inspection 197, 201, 202, 375, 392, 506
intensity 8, 9, 10, 11, 12, 15, 17, 30, 34, 36, 38, 39, 42, 51, 75, 80, 82, 85, 87, 88, 97, 102, 116, 117, 120, 146, 161, 193, 247, 263, 266, 269, 274, 275, 276, 277, 279, 283, 290, 291, 302, 310, 311, 313, 315, 319, 322, 323, 326, 327, 339, 340, 341, 342, 350, 357, 363, 365, 393, 394, 423, 424, 464, 465, 466, 469, 470, 471, 482, 483, 488, 491, 492, 497, 499, 501, 503, 507, 508, 509, 511, 512, 517, 518, 559, 562, 572
interference 38, 40, 41, 42, 52, 53, 56, 105, 106, 117, 120, 124, 246, 394, 412, 428, 459, 462, 491, 492, 515, 540
interferometer 18, 40, 41, 103
interferometry 60, 490
intermodal dispersion 108, 112
intrinsic semiconductor 70, 71, 72
inverse bremsstrahlung 289, 290, 320, 348, 382, 422
investment casting 479, 480
iron-carbon diagram 406, 431
irradiance 47, 117, 266, 314, 422

kerf 279, 280, 283, 284, 288, 289, 296, 297, 303, 305, 306, 329, 330, 497, 504, 509, 517, 522
kerf width 279, 280, 283, 288, 296, 303, 305, 306, 329, 330, 497, 509
keyhole 80, 139, 151, 158, 159, 161, 274, 308, 338, 339, 340, 341, 345–354, 365, 368, 369, 382, 384, 385, 386, 506, 508, 511, 516, 517
keyhole (deep penetration) welding 139, 151, 158, 338, 339, 345, 347–354
K factor 269, 277

laminated object manufacturing 461, 474–478, 481
lap joint 343, 344, 368, 379
laser
  argon ion (Ar^+) 32, 33, 60, 90, 463, 467, 483, 489, 490
  axial flow 65, 66, 90, 93
  carbon dioxide (C02) 57
  diode (semiconductor) laser 43, 69–80, 82, 85, 89, 90, 93, 273, 379, 381, 383, 488
  disk 55, 81–82, 89, 90
  excimer 62, 67–69, 90, 93, 273, 318, 412, 491, 493, 561
  fast axial flow 65, 66, 90
  fiber 55, 80, 84–90, 124, 273, 335, 336, 355, 356, 369, 379, 382, 383, 473
  helium-neon (He-Ne) 57
  heterojunction 76–77, 93
  homojunction 76, 93
  molecular 58, 61–69, 90
  neodymium glass (Nd:glass) 33, 56
  neodymium YAG (Nd:YAG) 85
  sealed tube 64–65
  slab 55, 80–81, 90
  slow axial flow 65–66, 93
  solid state 21, 22, 31, 34, 35, 37, 43, 55–57, 80, 81, 84, 89, 90, 92, 180, 192, 194, 335, 391, 413, 429, 501, 515, 517
  titanium sapphire 487
  transverse flow 64, 66–67, 90, 93, 267, 352
laser assisted machining 279, 323–326

laser assisted nanoimprint lithography 485, 493–494
laser beam analyzer 500, 501, 522
laser beam monitoring 497–504
laser cavity 13, 30, 31, 33, 44, 47, 50, 69, 85, 92, 508
laser cladding 171, 414, 415, 417
laser classification 562–563
laser cutting 17, 37, 38, 88, 116, 137, 139, 158, 175, 265, 279–330, 393, 486, 497, 504, 509, 510, 511
laser drilling 88, 137, 279, 308–318, 326, 327
laser forming 37, 151, 161, 176, 257, 437–450
laser generation 3–12
laser hazards
  eye 558–560
  non-beam 560–562
  radiation-related 557–560
  skin 563, 571
laser heat treatment 176, 249, 391, 392, 393, 400, 407, 411, 433
laser machining 324–326
laser physical vapor deposition 391, 419–420, 429
laser power 38, 80, 82, 266, 284, 288, 289, 291, 304, 307, 308, 322, 325, 326, 329, 330, 336, 337, 338, 341, 342, 345, 359, 367, 374, 375, 381, 414, 419, 420, 421, 429, 472, 498, 499, 508, 511, 516, 517
laser safety 557–571
laser safety officer 563, 571
laser shock peening 251, 391, 420–427, 429
laser surface melting 391, 413–414, 429
laser welded tailored blanks 335, 371–379
laser welding 37, 38, 79, 88, 120, 146, 156, 170, 201, 252, 265, 273, 335–386, 393, 413, 429, 497, 504, 509, 511, 517, 534
latent heat
  melting 135, 315, 351, 353
  vaporization 135, 312, 314, 315, 353
leakage 527, 531, 532, 548
least-squares miminum distance classification (classifier) 537–539
lens 22, 42, 43, 44, 79, 89, 95, 101, 102, 103, 104, 105, 116, 117, 118, 126, 129, 264, 265, 271, 275, 277, 278, 281, 284, 307, 310, 311, 322, 338, 339, 340, 341, 342, 369, 370, 380, 382, 385, 393, 394, 471, 558, 559, 560, 571
lifetime 8, 11, 24, 25, 26, 30, 31, 32, 43, 44, 50, 56, 60, 64, 72, 79
line broadening 26–29
linearly polarized 283, 289, 397, 399, 412
lineshape function 26, 27, 29, 50
line source 135, 141, 146, 150, 151, 161, 238, 239, 304
linewidth 15, 18, 21, 26, 27, 37, 38, 42, 44, 50, 52, 57, 65, 69, 129, 483
longitudinal (axial) mode 14–16, 18, 19, 33, 43, 44, 50, 51, 65
losses 9, 10, 18, 20, 24, 30, 31, 32, 35, 44, 47, 59, 61, 76, 77, 85, 106, 109, 114, 115, 116, 117, 119, 126, 129, 136, 194, 221, 263, 270, 272, 273, 276, 296, 314, 338, 356, 360, 363, 374, 394, 396, 399, 418, 419, 426, 489, 499, 532, 534, 536, 539, 546, 550, 560, 571

M^2 factor 267, 269, 276, 278, 503
manufacturing 17, 37, 52, 57, 89, 90, 114, 116, 119, 171, 263, 287, 437, 453–483, 485, 486, 487, 494, 540, 549
martensite 192, 193, 194, 195, 200, 201, 227, 238, 364, 384, 391, 400, 401, 402, 406–407, 410, 411, 413, 414, 432, 433
material dispersion 109, 111, 126
materials processing 55, 57, 62, 66, 69, 78, 79, 80, 87, 90, 93, 105, 116, 227, 263, 275, 557
maximum permissible exposure (MPE) 564, 565, 566, 568, 569
Maxwell's equations 157
mean square error 537, 538
medical device manufacturing 485
melting temperature (point) 57, 135, 140, 147, 150, 171, 178, 181, 196, 198, 206, 207, 209, 211, 212, 255, 287, 289, 296, 300, 303, 304, 307, 315, 316, 318, 319, 321, 329, 356, 358, 381, 386
Michelson interferometer 40, 41
micromachining 65, 84, 87, 279, 318–319, 324, 427, 428
microstructure 135, 136, 175–207, 209, 355, 360, 361, 364, 382, 391, 393, 397, 400–401, 413, 414, 419, 429, 439, 488, 497, 504
microwelding 335, 371, 383
minimum-distance classifier 539, 554
mirror 9, 10, 13, 15, 17, 18, 19, 20, 21, 31, 32, 40, 41, 43, 48–51, 56, 59, 65, 66, 67, 75, 82, 85, 92, 95, 101, 103, 105, 124, 126, 265, 266, 269, 271, 272, 275, 281, 349, 370, 385, 394, 395, 396, 500, 507–509
mode
  longitudinal (axial) 14–16, 18, 19, 33, 43, 44, 50, 51, 65
  transverse 14–16, 18, 19, 43, 48
mode-locking 33–35, 44, 51, 55, 82, 87
mode selection 18, 43
molecular lasers 58, 62

molten pool 135, 156–161, 171, 176, 179, 181, 182, 184, 185, 186, 187, 196, 197, 202, 203, 206, 207, 209, 217, 339, 345, 348, 351, 353, 354, 360, 365, 366, 367, 384, 385, 386, 438, 515, 516, 517, 518, 520, 522
momentum (Navier Stokes) equation 156–158, 217, 219, 220, 293, 296, 313, 350
monochromatic 8, 17, 18, 26, 35, 36–38, 40, 44, 52, 75, 76, 110, 111, 118, 245, 248, 263, 557
multi mode fiber 87, 88, 106, 107, 108, 111, 112, 113, 126, 128, 130
multi-photon absorption 320, 322, 327

nanobumps 485, 490–494
nanoholes 485, 490, 491, 492, 494
nanoimprint lithography 493–494
nanojets 490, 492, 494
nanotechnology 84, 427, 485–495
natural broadening 26, 27, 29, 44, 50
Navier Stokes (Momentum) equation 156–158, 217, 219, 220, 293, 296, 313, 350
neodymium-glass (Nd:glass) laser,see laser – Nd:glass)
neodymium YAG (Nd:YAG) laser, see laser – Nd:YAG)
neon 28, 58, 59
neural network 527, 534, 540–546, 550
nitrogen 76, 92, 105, 195, 196, 199, 284, 285, 295, 327, 340, 414, 515
nominal hazard zone (NHZ) 564, 569
nonmetals 65, 66, 88, 199, 269, 276, 299, 302, 318, 335
normal (Gaussian) distribution 16, 29, 135, 138, 151, 161, 263, 283, 401, 503, 536, 537
nozzle 69, 103, 272, 274, 281, 282, 287, 292, 297, 307, 308, 311, 318, 326, 327, 340, 341, 417, 509, 510
nucleation 175, 176, 177, 179, 180, 181, 182, 186–188, 196, 202, 203, 206, 207, 209, 222, 312, 408
numerical aperture (NA) 43, 85, 106, 107, 114, 126, 128, 129, 322, 471, 491

oligomers 469, 470
ophthalmology 488–489, 494
optical components 82, 272, 274, 276, 520
optical density 566, 567
optical fiber 76, 80, 82, 84, 85, 86, 89, 90, 105, 106, 107, 109, 116, 129, 130
optical integration 393–397
optical pumping 21–23, 57
oscillator 57, 82, 87
oxygen assist 137, 284, 285, 286, 287, 288, 298, 299, 300, 301, 303, 304, 327

part building 453, 461, 462–477, 481
pattern classification 527, 534–549
pattern recognition 527, 534–540, 542, 550
peak temperature 135, 136, 143, 146–148, 154, 161, 169, 171, 295, 304, 318, 350, 402, 403, 404, 410, 497, 516
pearlite 180, 193, 400, 402–406, 413, 432, 433
penetration 80, 156, 158, 159, 161, 175, 195, 201–203, 265, 267, 276, 311, 325, 326, 336, 337, 338, 340, 341, 342, 343, 345, 346, 347, 348, 359, 363, 365, 367, 381, 382, 394, 396, 399, 465, 467, 497, 504, 516, 517, 560, 571
percussion drilling 309, 310, 327
phase transformation 181, 192, 238, 325, 407, 411, 506
phonon 73, 88, 152, 153, 154, 514
photoacoustic effects 558, 569
photochemical effects 558
photocoagulation 488, 489
photodiode 511, 517, 520
photodynamic therapy 12, 488, 489
photoinitiators 469, 470
photon 3, 4, 5, 6, 9, 11, 12, 43, 69, 73, 114, 152, 153, 154, 280, 318, 320, 321, 349, 463, 469, 470, 471, 491, 493, 500, 514, 558
photopolymerization 469, 472
picosecond 33, 35, 77, 82, 152, 153, 318, 320, 427
piezoelectric transducer 504, 506, 507, 509, 523
Planck's constant 3, 38, 180
Planck's Law 7, 512
plane of incidence 96, 97, 101, 125, 129, 290, 399
plane strain 235–237, 252, 254
plane stress 235–237, 241, 252, 254
plasma
  absorption 273, 516
  density 323, 340, 511
  formation 289, 290, 310, 336, 339–340, 367, 516
plasma arc cutting 305, 306, 307
plastic burn analysis 503
plastics 79, 87, 88, 106, 195, 227, 228, 230, 231, 232, 233, 234, 235, 236, 238, 239, 240, 245, 251, 252, 258, 269, 319, 376, 377, 378, 379–380, 383, 422, 425, 440, 442, 443, 444, 445, 471, 472, 473, 476, 479, 503, 506, 562
p-n junction 73, 74, 75
point source 108, 139, 146, 294, 362
Poisson's ratio 233, 243, 257, 424
polarization 6, 10, 95, 98–101, 263, 274, 283, 289, 329, 335–338, 382, 412

polarized beam 283, 327, 338, 399
polymerization 463, 468–472, 478
population inversion 9, 10, 21, 24, 25, 26, 31, 32, 43, 44, 51, 56, 58, 59, 60, 63, 64, 67, 69, 72, 74, 75
porosity 80, 156, 175, 195–197, 199, 203, 355, 356, 358, 359, 368, 369, 382, 414, 417, 497, 504, 506, 527, 533
post-processing 453, 464, 472, 477–478
powder 187, 414, 415, 417, 418, 429, 462, 471–477, 481
power density 17, 43, 80, 82, 124, 125, 263–266, 275, 276, 277, 279, 283, 302, 310, 311, 314, 316, 325, 326, 327, 342, 360, 369, 429, 470, 557, 558, 569
p-polarized 290, 291, 399, 400
preheat 143, 146, 149, 161, 171, 193, 200, 201, 203, 250–251, 257, 302, 307, 355, 357, 358, 360, 361, 362, 363, 364, 382, 384, 397, 399–400, 418, 419, 432
principal strain 243, 244, 376
principal stress 234, 243, 244, 255
probability 5, 6, 24, 26, 47, 59, 69, 70, 73, 206, 533, 534, 535, 536, 538, 544, 550, 552, 553
probability density 534, 535, 536, 550, 552
process monitoring 497–524, 549
propagation 8, 12, 29, 96, 98, 100, 107, 108, 113, 122, 123, 192, 266, 269, 465, 466, 469, 470, 497, 498, 542, 543
pulse
  duration 29, 31, 33, 35, 38, 55, 57, 69, 77, 82, 84, 87, 90, 109, 152, 153, 154, 156, 267, 309, 310, 312, 313, 318, 319, 320, 321, 322, 327, 338, 412, 424, 425, 427, 428, 432, 470, 488, 489, 490, 491, 497, 558, 569
  repetition rate 29, 77, 82, 84, 93, 267, 283, 338, 412, 491, 497
  shape 267, 497
pulsed beam
  gated pulse 267
  hyperpulse 276
  superpulse 276, 489
pumping
  diode 22, 82, 87, 412
  electrical 21, 23–24, 43, 58
  flashlamp 21–22
  power 25
  rate 31–32
pyroelectric 498, 499, 500, 501, 502, 514

Q (quality) factor 30–31, 35, 44
Q-switching 31–33, 44, 82, 86, 267

quenching 59, 64, 65, 194, 195, 204, 238, 305, 391, 392, 413, 429

rapid prototyping 437, 448, 449, 453, 454, 479
rapid solidification 209, 213, 221–222, 413, 414, 429
rastering 393, 396, 397
rate equation 149, 363
Rayleigh
  range 20, 49, 104, 266
  scattering 115
  wave 507
recrystallization 171, 188, 191, 192, 194, 203, 251, 304
reflectance 166, 269, 299, 379, 380, 427
reflection 21, 22, 31, 81, 96, 97, 98, 102, 104, 106, 273, 274, 289, 314, 338, 349, 364, 403, 432, 433, 446, 448, 449, 450, 503
reflection coefficient (Reflectivity) 97, 98, 314, 403, 432, 433, 446, 449, 450, 503
refraction 17, 75, 106, 126, 471
refractive index 31, 35, 43, 59, 76, 84, 85, 87, 95, 98, 106, 107, 108, 110, 111, 115, 119, 128, 129, 322, 412
relaxation 153, 154, 156, 227, 240–244, 252, 324, 393, 426, 427, 428, 430
relaxation time 153, 427, 428
residual stress 228, 230, 231, 237, 238, 240, 241, 242, 249–250, 258, 259, 305, 411, 422, 425, 426
resin 463, 464, 466, 467
resistance 61, 182, 190, 192, 193, 194, 199, 241, 293, 299, 301, 325, 335, 356, 371, 372, 391, 403, 413, 414, 420, 429, 430, 432, 516
resonator
  concentric 20–21, 43
  confocal 18–20, 48, 49, 52
retina 88, 489, 494, 558–560, 569, 571
rosenthal equations 142
rotating hollow needle technique 500–501
rotating rod technique 499
rotation 33, 238, 239, 325, 500, 501
rotational energy level 61
roughness 283, 287, 289, 296, 297, 298, 303, 307, 422, 429, 486, 497
round trip gain 35
ruby laser 56

sampling 527, 529, 530, 531, 532, 544, 548, 550
sampling theorem 527, 531, 550
scanning 371, 393, 394, 396, 412, 413, 414, 429, 437, 446, 448, 449, 464, 465, 466, 468, 490, 491, 494, 500, 501, 502, 515

scattering 10, 88, 105, 106, 115, 116, 248, 274, 338, 349, 379
seam tracking 497, 517, 522
selective laser sintering 461, 472–473
semiconductor laser 23, 37, 43, 55, 69–80, 93
semiconductors 23, 37, 43, 55, 69–80, 93, 319, 320–322, 324, 411–412, 429
sensor 282, 463, 497–524, 527–554
sensor fusion 546
shielding gas 159, 199, 200, 263, 272, 276, 335, 338, 339–342, 348, 356, 375, 382, 393, 414, 415, 417, 418, 497, 511, 516
shock peening 251, 391, 420–427, 429, 432
shock wave 313, 314, 323, 391, 420, 422, 423, 425, 429, 511, 558
short-time Fourier transform 527, 547–548
silicon 71, 73, 79, 103, 171, 195, 196, 197, 199, 303, 356, 357, 411, 427, 428, 479, 491, 511
single mode fiber 87, 88, 106, 107, 108, 111, 113, 124
skin depth 152, 269, 276, 314, 321, 323
slicing 460, 461–462
slow axial flow laser 65, 66
small signal gain coefficient 9
solid modeling 454, 458–460
solid state laser 22, 31, 34, 35, 37, 55–57, 80, 81, 84, 89, 90, 92, 429
solidification rate 184, 185, 196, 203, 213
spatial coherence 38–40, 52
specific heat 135, 144, 147, 150, 153, 218, 271, 296, 303, 315, 316, 329, 351, 386, 404, 446, 499, 523
spectral analysis 508, 527
spectroscopy 61, 62
spiking 33, 547
spline 454–456, 458, 461, 482
s-polarized 290, 309
spontaneous emission 5–9, 26, 37, 38, 44, 47, 50
spot size 19, 20, 21, 48, 49, 266, 268, 275, 277, 283, 322, 337, 381, 382
stabilization 194, 195, 213, 214, 222, 223
stainless steel
  austenitic 193, 194, 195, 199, 203, 299
  ferritic 193, 194, 203, 299
  martensitic 193, 195, 203, 299
steels
  alloy 192, 193, 200, 203, 251, 299, 301, 325, 335, 355, 360, 361, 363, 370, 381, 382, 391, 392, 400, 401, 407, 408, 413, 414, 417, 419, 422, 423, 426, 428, 429, 439, 449, 450, 487
  eutectoid 148, 193, 402, 406, 410
  high carbon 192, 193, 200, 203, 361, 439
  low carbon 143, 191, 192–194, 201, 329, 359, 381, 382, 392, 413, 429, 449
  medium carbon 197
  plain carbon 148, 193, 299–301, 325, 329, 342, 355, 401, 402, 439
  pre-eutectoid 402, 410
Stefan-Boltzmann law 512, 513
stent 324, 485–488, 494
step index fiber 107, 108, 116, 128, 129
stereolithography 453, 463, 464, 465, 468, 472, 483
stimulated absorption 4, 5, 43
stimulated emission 3, 5, 6, 7, 9, 10, 43, 47, 73, 75
strength
  tensile 378
  yield 190, 228, 234, 238, 251, 376, 378, 424, 437, 440
striation 280, 287, 296–297, 307, 327, 486, 497, 504
substrate 124, 181, 182, 221, 311, 312, 318, 350, 411, 412, 413, 414, 415, 417, 418, 419, 420, 421, 429, 432, 490, 491, 493, 498, 515
supercooling 176, 177, 180, 206, 207, 212, 213, 214, 221, 222, 223
surface hardening 391
surface heat treatment 37, 67, 119, 269, 391–414, 429, 432, 437
surface melting 135, 156, 158, 161, 175, 181, 207, 221, 391, 393, 413–414, 429
surface modification 57, 90, 221, 263, 269, 391–413, 429, 473
surgery 57, 61, 62, 66, 486, 488, 494

tailored blanks 335, 371–379
TEM modes 16, 107
temperature
  profile 153, 155, 216
  gradient 80, 81, 143, 144, 161, 183, 184, 185, 186, 203, 210, 211, 213, 220, 222, 302, 312, 339, 352, 357, 360, 363, 438, 440–442, 444, 446, 448, 516
temperature gradient mechanism 440–441, 446, 448
temporal coherence 38, 39, 40–42
tensile strength 378
termination 469, 470
therapy 12, 488, 489
3D printing 454, 461, 473, 479
thermal coefficient of expansion 157, 350
thermal conductivity 57, 64, 136, 142, 143, 144, 147, 150, 161, 201, 267, 274, 276, 281, 284, 286, 299, 300, 301, 302, 303, 315, 325, 356, 357, 358, 386, 393, 404, 439, 440, 446, 503

thermal detectors 514
thermal diffusivity 137, 145, 147, 304, 314, 316, 351, 404
thermal expansion 227, 239, 302, 314, 358, 381, 411, 414, 440, 442, 443, 445, 446, 491, 493
thermal imaging 515
thermal lensing 80, 81, 82, 90, 104
thermal properties 165, 188, 286, 314, 400
thermal stress 135, 227–230, 251, 252, 357, 363, 414, 440, 442
thermal time constant 402, 403
thermalization 64, 153, 154, 156, 161
thermocouple 371, 498, 514
three-level system 24–26, 43, 86
threshold gain 9–10
threshold population inversion 31, 60
time constant 30, 49, 402, 403, 498
time-frequency analysis 546–549
titanium 187, 194, 195, 197, 199, 243, 244, 287, 300–301, 303, 305, 324, 325, 357, 361, 382, 426, 432, 439, 446, 449, 487
titanium alloys 300–301, 324, 325, 361, 439
titanium sapphire lasers 487
transform
  Fourier 30, 527–528, 530, 546, 547, 550
  orthogonal 546
transformation 79, 117, 120, 121, 122, 137, 138, 164, 169, 180, 181, 192, 193, 200, 205, 238, 275, 312, 325, 362, 363, 384, 391, 392, 400, 401, 402, 403, 404, 406, 407, 411, 413, 429, 432, 439, 454, 506, 527, 535, 537, 538, 539, 550, 553, 554
transformation hardening 275
transition 3, 5–8, 11, 15, 17, 18, 24, 25, 26, 29, 30, 43, 44, 46, 47, 50, 51, 52, 55, 56, 59, 60, 61, 62, 63, 64, 67, 69, 73, 76, 86, 92, 121, 185, 312, 341, 368, 400, 412
translation 61, 548
translational energy 61
transmittance 49, 380, 564, 566, 567
transverse electromagnetic mode (TEM) 16
transverse mode 14, 15, 16, 18, 19, 43, 48
traverse speed 143, 161, 186, 263, 282, 284, 327, 335, 336, 347, 382, 413, 416, 419
trepanning 241, 309, 310, 311, 327
two-level system 5, 24, 43, 46
two-photon
  absorption 10–12, 468, 470
  excitation 84
  polymerization 468–471
ultrashort pulse laser 82, 89, 135, 152, 161, 319, 322, 327, 488
ultraviolet detectors 504, 522
undercut 156, 175, 195, 202, 203, 359, 360, 365, 367, 497
upsetting mechanism 440, 441–442, 443, 448

valence band 69–75, 320
vaporization 135, 270, 279, 296, 308, 312, 313, 314, 315, 316, 318, 319, 320, 321, 322, 326, 339, 345, 346, 347, 349, 350, 351, 353, 355, 356, 358, 386, 427, 490
variance criterion 532, 550
velocity of light 3, 31
viscosity 157, 217, 219, 224, 267, 284, 286, 293, 294, 297, 298, 299, 301, 327, 386, 493
visibility 41, 197
vision sensing 517–522
volume fraction 215, 216, 219, 220, 222, 406, 407, 432, 433

waist size 20, 122, 266
warning signs 563, 569, 570
wavefront 38, 104, 117, 121, 122
waveguide dispersion 109, 112, 126
wavelength 3, 4, 7, 11, 12, 13, 14, 15, 17, 18, 22, 32, 36, 37, 39, 42, 43, 44, 46, 47, 48, 49, 50, 51, 52, 55, 56, 59, 61, 62, 64, 69, 76, 77, 80, 84, 85, 87, 88, 89, 90, 92, 95, 96, 98, 101, 105, 110, 111, 112, 114, 115, 116, 118, 119, 122, 123, 125, 126, 128, 129, 245, 248, 249, 257, 258, 263, 264–265, 269, 270, 275, 276, 277, 299, 310, 318, 319, 320, 324, 343, 369, 371, 381, 394, 397, 399, 412, 422, 428, 429, 463, 469, 470, 489, 490, 491, 492, 493, 499, 501, 502, 503, 507, 511, 512, 515, 518, 523, 587, 558, 559, 560, 561, 564, 565, 566, 567, 568, 569, 571
wavelet transform 548–549
wavenumber 14, 98, 107, 113
weld
  butt 17, 93, 171, 178, 266, 337, 360, 364, 382, 386
  lap 88, 355, 358
weld pool 178, 340, 341, 345, 348, 351, 354, 359, 360, 365, 382, 385, 386, 497, 504, 516, 519, 521
windowing 59, 65, 80, 117, 129, 272, 274, 418, 499, 530, 532, 547, 548, 550, 562, 564, 567

X-ray 3, 43, 227, 240, 245, 246, 247, 249, 252, 257, 258
xenon 22, 57, 60, 61
young's modulus 228, 233, 244, 249, 257, 324

yttrium aluminum garnet (YAG) 56

zinc selenide (ZnSe) 102, 104, 265, 271, 275
zone of partial melting 175, 188, 200, 202, 203, 413

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley’s ebook EULA.